C U R S O S E C O N G R E S O S

VII Conference of the Spanish Society for Logic, Methodology and Philosophy of Science

VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España

Santiago de Compostela
18-20 July 2012
18-20 de julio de 2012

Edited by
Concha Martínez Vidal
José L. Falguera
José M. Sagüillo
Víctor M. Verdejo
M. Pereira-Fariña

UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA
publicacións

CURSOS E CONGRESOS
DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
Nº. 217

VII Conference of the Spanish Society for Logic, Methodology and Philosphy of Science

VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España

VII Conference of the Spanish Society for Logic, Methodology and Philosphy of Science

VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España

Santiago de Compostela, Spain, 18-20 July 2012

Santiago de Compostela, España, 18-20 de julio de 2012

Edited by

Concha Martínez Vidal

José L. Falguera

José M. Sagüillo

Víctor M. Verdejo

M. Pereira-Fariña

2012

Universidade de Santiago de Compostela

Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España. Congreso (7º. 2012. Santiago de Compostela)
VII Conference of the Spanish Society for Logic, Methodology and Philosphy of Science, Santiago de Compostela, Spain, 18-20 July 2012 = VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España, Santiago de Compostela, España, 18-20 de julio de 2012 [Recurso de internet] / edited by Concha Martínez Vidal, José L. Falguera, José M. Sagüillo, Víctor M. Verdejo, M. Pereira-Fariña. — Santiago de Compostela : Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico, 2012
(Cursos e congresos da Universidade de Santiago de Compostela ; 217)
Formato PDF
Requisitos do sistema : Adobe Acrobat Reader
Modo de acceso : Internet repositorio Minerva, http://dspace.usc.es/ (15-07-12)

1. Ciencias — Metodoloxía — Congresos I. Martínez Vidal, Concha, ed. lit. II. Falguera, José L., ed. lit. III. Sagüillo, José M., ed. lit. IV. Verdejo, Víctor M., ed. lit. V. Pereira-Fariña, M., ed. lit. VI. Universidade de Santiago de Compostela. Servizo de Publicacións e Intercambio Científico, ed.

001.8:061.3(461.11 Santiago de Compostela)

http://hdl.handle.net/10347/5853

This work is under a Creative Commons BY-NC-SA 3.0 license.
Any form of reproduction, distribution, public communication or transformation of this work not included under the Creative Commons BY-NC-SA 3.0 license can only be carried out with the express authorization of the proprietors, save where otherwise provided by the law.
You can access the full text of the license by clicking on the following link:
http://creativecommons.org/licenses/by-nc-sa/3.0/

Esta obra se encuentra bajo una licencia Creative Commons BY-NC-SA 3.0.
Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra no incluida en la licencia Creative Commons BY-NC-SA 3.0 solo puede ser realizada con la autorización expresa de los titulares, salvo excepción prevista por la ley.
Puede acceder Vd. al texto completo de la licencia haciendo clic en este enlace:
http://creativecommons.org/licenses/by-nc-sa/3.0/es/legalcode.es

© Universidade de Santiago de Compostela, 2012

http://hdl.handle.net/10347/5853

Maquetación
Emanuela Saladini

Edita
Servizo de Publicacións e Intercambio Científico
Campus Vida
15782 Santiago de Compostela
usc.es/publicacions

INDEX / ÍNDICE

Preface XIII

Prefacio XV

Premio para Jóvenes Investigadores XVI

Contributed papers / Comunicaciones 1

1. LOGIC, HISTORY AND PHILOSOPHY OF LOGIC / LÓGICA, HISTORIA Y FILOSOFÍA DE LA LÓGICA 2

MARTIN ANDOR: Inferentialism and Expressivism about Logical Vocabulary in Frege 3

JOHN CORCORAN and JOSÉ SAGÜILLO: Information-Theoretic Approaches to Logical Relations 11

MANUEL CRESCENCIO MORENO GOMEZ: La Historia de la Lógica como la Historia de los Argumentos Ontológicos 16

JOAQUÍN DÍAZ BOILS: La Jerarquía de Grzegorzcyk en Teoría de Categorías 27

EMILIO GARCIA BUENDIA: Abducción y Derecho 46

PIERDANIELE GIARETTA and SILVIA GAIO: Gödel's Syntactical Conventions 53

IGNACIO HERNÁNDEZ-ANTÓN, ENRIQUE SARRIÓN-MORILLO and FERNANDO SOLER-TOSCANO: Abducción y Semántica de Teoría de Juegos 60

JOHN KEARNS: First-person Logical Theories and Third-Person Theories 67

JOAN ROSELLÓ: Incompleteness, Undecidability and Platonism in the Work of Kurt Gödel 74

ENRIQUE SARRIÓN-MORILLO, IGNACIO HERNÁNDEZ-ANTÓN and ÁNGEL NEPOMUCENO-FERNÁNDEZ: Tratamiento multimodal de contextos 81

JOSÉ PEDRO UBEDA RIVES: Una lógica epistémica bidimensional 88

MARGARITA VÁZQUEZ: Tiempo y puntos de vista 95

2. PHILOSOPHY OF LANGUAGE, PHILOSOPHY OF MIND, EPISTEMOLOGY / FILOSOFÍA DEL LENGUAJE, FILOSOFÍA DE LA MENTE, EPISTEMOLOGÍA 101

JUAN JOSÉ ACERO FERNÁNDEZ and NEFTALÍ VILLANUEVA FERNÁNDEZ: Wittgenstein's Anti-descriptivism 102

SILVIA ANDRÉS BALSERA: Unidad conceptual versus multiplicidad de grupos de deferencia 109

MARC ARTIGA GALINDO: The Singular Thought Strategy and the Content of Perception 114

ANTONIO BENÍTEZ and EMILIO GARCÍA BUENDÍA: El Modelo Animat en la Inteligencia Artificial bio-inspirada 122

EDSON BERND, ANTONIO GAITÁN-TORRES and MANUEL DE PINEDO-GARCIA: Two Kinds of Second Person Authority 132

ANTONIO BLANCO SALGUEIRO: ¿Fue Whorf un neowhorfiano? 139

DAVID BORDONABA and NEFTALÍ VILLANUEVA: Lo que se dice y "decir lo mismo" 147

FERNANDO BRONCANO-BERROCAL: Luck and Fortune, or Why Knowledge Requires Cognitive Control 155

GEMMA CELESTINO FERNÁNDEZ: It Is Not the Case that the Alleged Golden Mountain Exists 162

CRISTINA CORREDOR: Communicative Intentions and Joint Commitments in Speech Acts 170

ANTONIO DUARTE: El arte del diagnóstico: falacias irónicas 177

VÍCTOR FERNÁNDEZ: Metarepresentación y Dualismo en la Lectura de Mentes 184

LUIS FERNÁNDEZ MORENO: Nombres propios y términos de género natural: un examen de dos supuestas similitudes 191

EKAIN GARMENDIA: Boghossian's Argument from Inference and Identity Belief Ascriptions 198

HEIMIR GEIRSSON: A Tale of Two Mary's 205

MARIO GÓMEZ-TORRENTE: On Quotation and Semantic Structure 212

EDUARDO GONZÁLEZ DE LUNA: Una noción dinámica de la representación mental en el marco de un realismo natural 216

JAVIER GONZÁLEZ DE PRADO SALAS: Inferentialism, Metaphor and Context 223

MANUEL HERAS-ESCRIBANO and MANUEL DE PINEDO-GARCÍA: Dispositions, Affordances and Mental Intentionality 230

MARTA JORBA: Concept-aspects and the Specification of Cognitive Phenomenology 236

MANUEL LIZ: Decisiones epistémicas y los límites del naturalismo 241

MANUEL LIZ: "Yo soy George Kaplan". Tres tesis sobre la rigidez semántica 246

MIREIA LOPEZ AMO: ¿Es la tesis de la carga teórica de la observación incompatible con el dogmatismo? 253

JANUSZ MACIASZEK: Metaphor: Meaning, Interpretation and Causation 261

MANUEL PÉREZ OTERO: Argumentación racional sin transmisión de la justificación 269

PABLO RYCHTER: Presentism, Truthmaking, and Laws of Nature 274

UXÍA RIVAS MONROY: Creencia, conocimiento y verdad: el enfoque lógico-semántico de Frege y Russell y el pragmatista de Peirce y James 280

MARÍA ESPERANZA RODRÍGUEZ ZARAGOZA: ¿Qué hay detrás de un argumento? La distinción objetividad-subjetividad mostrada a través de la distinción entre desacuerdo perdido y desacuerdo legítimo 287

SERGI ROSELL: Voluntary Belief on a Strictly Epistemic Basis? 295

JONATHAN SUROVELL: Carnap's Principle of Tolerance: Rejection or Neglect of Language 300

BARBARA TRYBULEC: Is Naturalism Normative? Some Problems with Naturalizing Epistemic Normativity 308

VÍCTOR M. VERDEJO: Unbelievably Good Norms 316

IGNACIO VICARIO: El rompecabezas de la atribución de creencias 324

JAVIER VILANOVA: La percepción del aspecto en las *Investigaciones Filosóficas* 331

3. PHILOSOPHY AND METHODOLOGY OF SCIENCE / FILOSOFÍA Y METODOLOGÍA DE LA CIENCIA 339

ALBA AMILBURU: Natural Kind Thinking in Philosophy of Science: a Meta-theoretical Analysis 340

LORENZO BARAVALLE: Extensión del dominio de la teoría de la selección natural: enfoques sincrónicos y diacrónicos del cambio cultural 343

MARÍA CAAMAÑO: Las anomalías no refutatorias y el alcance del dominio teórico 350

EDUARDO CASTRO: Induction, Laws of Nature and Inference to the Best Explanation 358

GUSTAVO CEVOLANI, VINCENZO CRUPI and ROBERTO FESTA: A Verisimilitudinarian Analysis of the Linda paradox 366

GUSTAVO CEVOLANI and LUCA TAMBOLO: Underdetermination, Realism, and Truth Approximation 374

JUAN J. COLOMINA: The Feasibility of Determinables and its Relation to the Scientific Image 382

TIM DE CRAECKER: Causal Explanation in Theories of Probabilistic Causation 389

XAVIER DE DONATO-RODRIGUEZ: Actitudes cognitivas: el caso de la asunción, particularmente en ciencia 396

ALEX DÍAZ: The Müller-Lyer Visual Illusion: Tricking the Brain or Tricking the Scientist? 403

JOSÉ L. FALGUERA: De los géneros constituidos de Kuhn a los objetos abstractos de Zalta 409

M. J. GARCÍA-ENCINAS: Reading *De Interpretatione* 9 417

ANDONI IBARRA and HANNOT RODRÍGUEZ: Gobernanza de riesgos de la nanomedicina: el caso europeo 424

MARIA JIMENEZ-BUEDO: Experiments in the Social Sciences: Rethinking the Hawthorne Effect 432

VÍCTOR JOSÉ LUQUE MARTÍN: La Explicación en Biología: entre Leyes y Modelos Explicativos 437

ALFREDO MARCOS: Implicaciones filosóficas de la biología post-genómica 444

IGNACIO MARTÍNEZ DE LEJARZA: Medidas probabilistas de apoyo evidencial: ¿pluralidad de medidas o pluralidad de relaciones? 451

THOMAS MEIER: Un marco conceptual para el realismo estructural 465

ELSA MURO and MARÍA CEREZO: Información biomolecular con contenido semántico. Una perspectiva sistémica 473

VIOREL PÂSLARU: Conceptions of Mechanisms and Insensitivity of Causation 481

DIANA LUZ RABINOVICH: Medición en Psicología. Ponderación de las concepciones clásicas 489

ANDRÉS RIVADULLA: Theoretical Preduction and Computational Scientific Discovery 497

CRISTIAN SABORIDO: La problemática naturalización de la normatividad natural. Organización y malfuncionalidad biológica 505

MARGARITA SANTANA: La discusión del positivismo lógico en la JAE: Blas Cabrera y Julio Palacios 513

IÑAKI SAN PEDRO: Violation of Measurement Independence without Conspiracy 520

IULIAN TOADER: An Outline of Weylean Skepticism 529

NICOLÁS VENTURELLI: Mecanismos, sistemas dinámicos y el Santo Grial de la explicación científico-cognitiva 536

SUSAN VINEBERG: Indispensability and Mathematical Explanation 544

ZENAIDA YANES ABREU: Verdades significantes: o cómo incluir lo político en la ciencia 552

EDUARDO ZUBIA: La lógica de las imágenes: retórica y argumentación en las prácticas científicas representacionales 559

4. HISTORY OF SCIENCE / HISTORIA DE LA CIENCIA 567

CARLOS ALBERTO CARDONA Panofsky: el conflicto entre la perspectiva lineal y la perspectiva angular 568

PABLO RUIZ DE OLANO: Apunte acerca de la filosofía de la ciencia de Blas Cabrera. El papel de Duhem in su defensa de la relatividad 579

ALBERT SOLÉ: Three-dimensionalist Bohmian Mechanics: Does Newton Defeat Aristotle? 587

Mª JOSÉ TACORONTE DOMÍNGUEZ: El planteamiento sobre la mujer: positivistas, krausistas y católicos 595

5. SCIENCE, TECHNOLOGY AND SOCIETY / CIENCIA, TECNOLOGÍA Y SOCIEDAD 603

IRENE DÍAZ GARCÍA: La nueva ley de la ciencia en España: una aproximación crítica al contrato social subyacente 604

MANUELA FERNÁNDEZ PINTO: Aprendiendo de la ignorancia: la agnotología como nuevo enfoque en filosofía de la ciencia 611

BELEN LASPRA: Cultura científica y museos de ciencia: una historia paralela 618

JOSÉ LUIS LUJÁN, OLIVER TODT and JUAN BAUTISTA BENGOETXEA: Las reglas de inferencia como hipótesis empíricas. La información mecanística en la evaluación de riesgos 626

AMPARO ROMERO: El tecnofeminismo de Judy Wajcman. Alcance teórico y político 631

ANA ROMERO DE PABLOS: Las patentes, nuevas fuentes para la historia de la ciencia y los estudios CTS 639

MIGUEL ZAPATA: Los imaginarios colectivos en la representación social de la ciencia 642

Symposia / Simposios 649

1. LOGIC, HISTORY AND PHILOSOPHY OF LOGIC / LÓGICA, HISTORIA Y FILOSOFÍA DE LA LÓGICA 650

JOSÉ FERREIRÓS and MARÍA JOSÉ FRÁPOLLI (coord.), CONCEPCIÓN MARTÍNEZ VIDAL, JOSÉ MIGUEL SAGÜILLO, VALERIA GIARDINO, RAYMUNDO MORADO: Prospects for Practice-based Philosophy of Logic 651

ALEJANDRO SOBRINO and MARTÍN PEREIRA FARIÑA (coord.), SANTIAGO FERNÁNDEZ, MARÍA NAVARRO: II Seminario Alfredo Deaño sobre Razonamiento Ordinario 689

LUIS VEGA-REÑÓN (coord.), BEGOÑA CARRASCAL, XAVIER DE DONATO, HUBERT MARRAUD, RAYMUNDO MORADO, PAULA OLMOS GÓMEZ: Teoría de la argumentación: perspectivas y problemas 726

LUIS VEGA REÑÓN (coord.), ENRIQUE ALONSO Y MIGUEL VIVANCO, LILIAN BERMEJO LUQUE, JOSÉ M. SAGÜILLO: La argumentación en la esfera pública del discurso 763

2. PHILOSOPHY OF LANGUAGE, PHILOSOPHY OF MIND, EPISTEMOLOGY / FILOSOFÍA DEL LENGUAJE, FILOSOFÍA DE LA MENTE, EPISTEMOLOGÍA 784

JUAN COLOMINA and MANUEL LIZ (COORD.), FERNANDO CHARRO, DAVID P. CHICO MARGARITA VÁZQUEZ: Puntos de vista 785

JESÚS VEGA ENCABO and VÍCTOR M. VERDEJO (COORD.), LAURA ORTEGA CANO, MARÍA MUÑOZ-SERRANO: Semantics and Epistemology of Artifactual Kinds 801

3. PHILOSOPHY AND METHODOLOGY OF SCIENCE / FILOSOFÍA Y METODOLOGÍA DE LA CIENCIA 831

ANTONIO DIÉGUEZ (coord.), PASCUAL MARTÍNEZ-FREIRE, ALVARO MORENO, JON UMEREZ, ARANTZA ETXEBERRIA: Mecanismos y modelos en biología y en ciencias cognitivas 832

RODOLFO GAETA (coord.), NELIDA GENTILE, SUSANA LUCERO: Realismo y antirrealismo científicos: el caso de las leyes 868

4. SCIENCE, TECHNOLOGY AND SOCIETY / CIENCIA, TECNOLOGÍA Y SOCIEDAD **890**

EULALIA PÉREZ SEDEÑO and MARÍA JESÚS SANTESMASES (coord.), ANA TOLEDO, CARMEN ROMERO, MARÍA GONZÁLEZ AGUADO, LORENA RUIZ MARCOS: Cartografías del cuerpo: biopolíticas de la ciencia y la tecnología I Simposio 891

MARÍA JESÚS SANTESMASES and EULALIA PÉREZ SEDEÑO (coord.), MARÍA JOSÉ MIRANDA, ESTHER ORTEGA, SILVIA GARCÍA DAUDER , CARMEN ROMERO: Cuerpos en tránsito: biopolíticas de la ciencia y la tecnología II Simposio 918

PREFACE

Preface

This book comprises the works accepted to be presented at the VII Congress of the Society of Logic, Methodology and Philosophy of the Science in Spain (SLMFCE), celebrated in Santiago de Compostela, July 18th-20th, 2012. That is to say, it contains the contributions and symposia accepted for the different sections of the Congress. The volume incorporates a wide and representative sample of the research done in our country in the fields of logic, philosophy of science, philosophy of language, philosophy of mind, epistemology, history of science, and in science, technology and society. Besides, more than a third of the works received came from other parts of the world (Australia, United States and several countries in Latin America and Europe). This confirms the tendency to internationalization of the conferences of the SLMFCE, a tendency already initiated in previous editions. This is a good sign of the potential of these events to open and preserve communication between researchers of different countries.

All the proposals received were subjected to a process of blind refereeing. This process involved a good number of scholars from different countries and research interests as the composition of the Scientific Committee shows. The global rate of acceptance has been 74 percent, although this rate depended on the section and was lower in the sections of the congress with a larger number of contributions (Philosophy and Methodology of Science, and Philosophy of Language, Philosophy of Mind, Epistemology).

A novelty in this seventh edition of the Congress of the Society for Logic, Methodology and Philosophy of Science in Spain was the celebration of the first edition of a series of plenary lectures, the LULLIUS LECTURES. These lectures constitute an excellent tool by which distinguished philosophers can present their latest results to our Society. In this first edition the LULLIUS LECTURES have been honored with the presence of Professor Philip Kitcher, from the University of Columbia.

Besides, and to encourage the work of young researchers, the SLMFCE summoned the second edition of the prize “FINANCIAL SUPPORT FOR YOUNG RESEARCHERS”. In this edition, there were 17 applications.

The editors would like to thank the institutions whose financial support made this event possible: Ministerio de Economía y Competitividad, Universidade of Santiago de Compostela, and the Society for Logic, Methodology and Philosophy of Science in Spain.

The editors would also like to show their gratitude to the rest of members of the Organizing Committee, and to those who collaborated in the refereeing process, including indeed the Scientific Committee. Finally, we would like to thank the members of the board of the Society for Logic, Methodology and Philosophy of Science in Spain, for their advice and constant support, and all participants in the Conference for their contribution to its success.

The editors
Concha Martínez Vidal, José L. Falguera, José M. Sagüillo, Víctor M. Verdejo and M. Pereira-Fariña

Santiago de Compostela, July 10th 2012

PREFACIO

Prefacio

Este libro incluye los trabajos aceptados para ser presentados en el VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España (SLMFCE), celebrado en Santiago de Compostela del 18 al 20 de julio de 2012. Es decir, contiene las contribuciones aceptadas para las secciones del Congreso y las que forman parte de los simposios cuya celebración en el seno del Congreso fue aprobada. El volumen incorpora una muestra amplia y representativa, del trabajo que se realiza en nuestro país en el terreno de la lógica, la filosofía de la ciencia, la filosofía del lenguaje, la filosofía de la mente, la epistemología, la historia de la ciencia, y en los estudios sobre ciencia, tecnología y sociedad. Además, más de un tercio de los trabajos recibidos proceden de otras partes del mundo (Australia, Estados Unidos y varios países de América Latina y de Europa). Esto confirma la tendencia a la internacionalización de los congresos de la SLMFCE, tendencia ya iniciada en ediciones anteriores y que es un buen indicador del potencial de estos eventos para abrir y preservar cauces de comunicación entre investigadores de diferentes países.

Todas las propuestas recibidas fueron sometidas a evaluación "ciega". Este proceso involucró un buen número de académicos e investigadores de distintos países y especialidades tal y como se refleja en la composición del Comité Científico. El porcentaje de aceptación medio ha sido del 74 por ciento, aunque no todas las secciones tuvieron la misma tasa de aceptación, siendo ésta más baja en las secciones del congreso con un número de contribuciones más numeroso (filosofía y metodología de la ciencia, filosofía del lenguaje, filosofía de la mente, epistemología).

Como novedad, este congreso de la Sociedad incorporó la celebración de la serie "CONFERENCIAS LULLIUS" de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España. Esta serie de conferencias contará con la presencia de filósofos de prestigio que den a conocer sus últimos trabajos en el seno de nuestra Sociedad. En su primera edición las conferencias corrieron a cargo del profesor de la Universidad de

Columbia (Estados Unidos) Philip Kitcher.

Además, y con el fin de seguir incentivando el trabajo de jóvenes investigadores, la SLMFCE convocó la segunda edición de los PREMIOS/BOLSAS DE VIAJE PARA JÓVENES INVESTIGADORES. En esta edición, se recibieron 17 solicitudes.

En el capítulo de los agradecimientos, los editores no podemos dejar de referirnos a las instituciones sin cuya ayuda financiera no hubiera sido posible llevar a cabo este proyecto: Ministerio de Economía y Competitividad, Universidade de Santiago de Compostela, y Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España.

Queremos manifestar también nuestro agradecimiento al resto de miembros del comité organizador, y a quienes han colaborado en el proceso de revisión de las contribuciones recibidas, incluyendo, desde luego, al Comité Científico del congreso. Por último, quede constancia de nuestro reconocimiento a la directiva de la SLMFCE, por su asesoramiento y apoyo constantes, así como a todos los autores de las contribuciones y a los coordinadores de los diversos simposios.

Los editores

Concha Martínez Vidal, José L. Falguera, José M. Sagüillo, Victor M. Verdejo y M. Pereira-Fariña

Santiago de Compostela, 10 de julio de 2012

Premio para Jóvenes Investigadores
VII Congreso de la SLMFCE

Con motivo de la VII edición del congreso de la Sociedad de Lógica, Filosofía y Metodología de la Ciencia, y con el fin de incentivar el trabajo de jóvenes investigadores, la Sociedad decidió convocar y financiar la segunda edición de los premios, consistentes en bolsas de viaje, para las mejores contribuciones al congreso a cargo de investigadores jóvenes.

Las ponencias ganadoras fueron:

Marc Artiga Galindo - 'The Singular Thought Strategy'

Cristian Saborido - 'La problemática naturalización de la normatividad natural. Organización y malfuncionalidad biológica'

Juan J. Colomina - 'The Feasibility of Determinables and Its Relation to the Scientific Image'

Accésits

Víctor José Luque Martín - 'La explicación en biología: entre leyes y modelos explicativos'

Alex Díaz - 'The Müller-Lyer Visual Illusion: Tricking the Brain or Tricking the Scientist?'

Marta Jorba - 'Concept-aspects and the Specification of Cognitive Phenomenology'

Contributed papers / Comunicaciones

LOGIC, HISTORY AND PHILOSOPHY OF LOGIC

LÓGICA, HISTORIA Y FILOSOFÍA DE LA LÓGICA

INFERENTIALISM AND EXPRESSIVISM ABOUT LOGICAL VOCABULARY IN FREGE

MARTIN ANDOR
UGR

ABSTRACT: In (Brandom 1994) and (Brandom 2001), Robert Brandom controversially characterizes young Frege's philosophy of language as inferentialist. In Brandom's view, the kind of inferentialism present in the *Begriffsschrift* is based on the following three theses (Brandom 2001, 61):

(1) Conceptual content is to be understood in terms of role in reasoning.
(2) Such reasoning is not merely logical in nature.
(3) Logical vocabulary makes explicit the implicit material inferential roles of non-logical expressions.

Under Brandom's particular interpretation, (1) amounts to what may be called 'minimal global inferentialism' ('global' meaning here that no aspect of conceptual content should be accounted for in non-inferential terms, for instance by resorting to different semantics for some subsets of the lexicon). The second tenet is required to bestow plausibility on the first one, as logical inference is generally considered too restrictive to underpin a global semantic theory for natural language. The third tenet requires the first two, but is not necessitated by them.

In this paper I argue that none of these theses is actually held by Frege. I proceed by first showing that the arguments Brandom supplies for attributing the relevant interpretation of (1) to Frege are unsatisfactory. While this undercutting does not suffice by itself to reject this attribution, I make a strong case for interpreting 'inference' ('*Schluss*') in *Begriffsschrift* always as 'logical inference'. As noted above, the absence of (2) renders Brandom's (1) extremely implausible; however, I argue that one can attribute (1) to Frege under an interpretation that doesn't require (2). Without Brandom's reading of (1) and (2), (3) falls. I show why Brandom's reasons for attributing (3) to Frege don't stand closer scrutiny.

KEY WORDS: Frege, Brandom, inferentialism, expressivism, conceptual content

1. Sätze, propositions and sentences

Before proceeding to question Brandom's interpretation of young Frege's philosophy of language, a word is in order about the conceptual background supporting the theoretical system of each of the two philosophers.

We will follow the general practice of translating Frege's '*Satz*' by 'proposition', which shows the same kind of ambiguity between referring to some content and the linguistic device used for expressing this content (Dummett 1991, 23). In *Begriffsschrift*, Frege often treats the notions of proposition and judgeable content (*Beurteilbarer Inhalt*) as equivalent. However, two different

propositions in this sense can share their conceptual content (*Begrifflicher Inhalt*), which constitutes the inferentially relevant part of their judgeable content, as we will see in what follows.

On the other hand, it is arguably the notion of *claim* that has pride of place within Brandom's account of human linguistic practice. Pragmatically, claims amount to acknowledgments of discursive commitments. From a semantic point of view, claims are bearers of content that is conferred upon them by the rules that situate the corresponding commitment in the game of giving and asking for reasons. These contents are expressed explicitly by declarative sentences, which are syntactic entities (Brandom 1994, 91).

Since Brandom tries to argue that Frege maintained a similar position about the origin of conceptual content as he does himself, he tends to treat Frege's *Satz* as a syntactic item. Although I think this is wrong, I will follow Brandom's construal of Frege's conceptual framework all the way to the point where it is no longer viable, so as to avoid begging the question. This article may thus be regarded as a *reductio* of such a construal.

2. Minimal global inferentialism

Let us recall Brandom's first criterion for inferentialism:

(1) Conceptual content should be understood in terms of role in reasoning.

What are Brandom's arguments for attributing (1) to Frege? Brandom quotes the definition of conceptual content from the §3 of *Begriffsschrift*:

> "There are two ways in which the content of two judgments may differ; it may, or it may not, be the case that all inferences that can be drawn from the first judgment when combined with certain other ones can always also be drawn from the second when combined with the same other judgments." (Brandom 1994, 95)

If two propositions only differ in the first sense, Frege defines their conceptual content as the intersection of their contents. Conceptual content is the only part of semantic import of propositions relevant for the purposes of *Begriffsschrift*.

Putting aside the question of the nature of this inference for a moment, there is certainly a sense in which these remarks by Frege imply (1). Indeed, if it is in view of their inferential behaviour how we demarcate the conceptual content of our propositions, then the latter can properly be said to be 'understood' in terms of the former. However, there are two possible readings of Frege's passage, (and thus of (1) itself) in play – a weak one, under which it amounts to a definition of a technical notion of conceptual content in terms of the previously familiar notions of content and

inference; and a strong one, which interprets it as an account of how our sentences *acquire* their conceptual content in the first place. The former is the received view among Frege's scholars (Dummett 1991, 16; Beaney 1996, 56), whereas Brandom opts for the strong reading:

> "The fundamental semantic assignment of conceptual content to judgments is derived from the ultimately pragmatic notion of *correctness of inference*." (Brandom 1994, 96)

Also, when Brandom speaks about inferences in Begriffsschrift, he refers to them as "content-conferring material inferences" (Brandom 2001, 58).

Without this bold departure from the tradition, there would hardly be any good reason for calling Frege an inferentialist. Under the weak interpretation of the above passage, Frege's definition of conceptual content is also compatible with a referential account of content and a logicist notion of inference. And this is exactly what Carnap does in *The Logical Syntax of Language*, to which Brandom himself refers as bearing the legacy of Frege immediately after discussing Frege's definition of conceptual content (Brandom 1994, 96). As Carnap witnesses, simply conflating semantically the propositions that yield the same sets of consequences is not enough for Brandom's type of inferentialism. When Brandom characterizes inferentialism in the previous sections of the two books, he stresses the requirement of an independent previous account of inferential relations among propositional attitudes:

> "States and acts acquire content by being caught up in inferences, as premises and conclusions." (Brandom 2001, 46)

In turn, conceptual content is conferred upon sentences that are typically used in the expression of such states or in the performance of such acts (Brandom 1994, 91).

Now that we have clarified the issue at hand, it is time to assess the case Brandom makes for his stronger interpretation of Frege. And there are many reasons to regard it with suspicion.

First, it only relies on the brief passage reproduced above. It is perplexing that Frege should devote so little space to such a novel and fundamental semantic insight. If Frege had held this view, then its careful presentation should have been one of the main pillars of the *Begriffsschrift*, as Brandom indeed claims it is (Brandom 2001, 50). However, Brandom's claim is misleading: even if his interpretation were right, it is simply not true that the *Begriffsschrift* "takes as its aim the explication of 'conceptual content'". Fundamental though the notion is, it only prepares the ground for the main developments of the *Begriffsschrift*'s project.

For the weak interpretation, of course, there is no mystery involved. Since it attributes to

Frege an unanalysed notion of content and treats the notion of conceptual content as a technical one, not an intuitive one which Frege is trying to clarify, his brief comments on the topic can be considered sufficient for his purposes.

Second, there are other passages from Frege's early period that seem to problematize the identification of conceptual content and inferential role:

> Jene Unterscheidungen […] betreffen nun nach meiner Auffassung nicht den Inhalt des Urtheils, sondern die Berechtigung der Urtheilsfällung. (Frege 1884, §3)

In this case, justification yields the fundamental inferential antecedents of a proposition, but Frege nevertheless contends that the content of a judgment can often be gained without knowledge of them (ibid.). However, that goes directly against Brandom's construal of both *Begriffsschrift* and *Grundlagen* (Brandom 1994, 96).

Third, under the weak interpretation of Frege's definition, we seem to be in a better position to account for Frege's intellectual evolution from his early to his middle period. From *Über Sinn und Bedeutung* onwards, the notion of conceptual content splits into sense and reference. Brandom's interpretation of *Begriffsschrift* renders this innovation as a complete retraction from Frege's previously held positions. Again, although we have many examples of philosophers whose thought underwent profound transformations throughout their life, it would be striking for Frege never to address this huge shift in semantics in his work. When he does refer to it, he explains it as a natural elaboration of his previous framework (Frege 1892, 198). Dummett argues that middle and late Frege sometimes avoided explicit discussion of his early theses in view of the innovations he introduced, because he regarded the relation as obvious (Dummett 1991, p.2). However, such explanation would be excluded in case of a shift from inferential to referential semantics.

3. Material versus logical inference

As adumbrated in the introduction, I don't regard the arguments presented in the previous section as conclusive, although they certainly cast a shadow of doubt on Brandom's inferentialist construal of *Begriffsschrift*. In this section I will show how Brandom's reading of (1) in any reasonable semantic theory necessitates (2), after which I will argue that Frege had *logical* inference in mind when defining conceptual content. The upshot is that Frege was either very naïve about his semantics or did not hold a strong version of (1) in the first place.

Why would a defender of minimal global inferentialism want to embrace something like (2)? The reason is that if you are serious about proprieties of inference bestowing conceptual content on

your expressions, logical consequence simply is not rich enough for the task. Two sentences that differ only in one non-logical item of the same grammatical category will yield structurally similar sets of consequences, in which the actual semantic difference between the two non-logical items will never show. We will never be able to conclude whether they are synonyms, antonyms or neither.

A global inferentialist would, therefore, need a weaker notion of inference. Brandom calls this kind of inference *material*, as opposed to *formal*. One of Brandom's examples is the inference from "Lightning is seen now" to "Thunder will be heard soon" (Brandom 1994, 98), but just how weak the notion of inference should be is open to debate. We will now show that even a conservative notion of material inference would be at odds with Frege's logicist project.

When Brandom quotes Frege's presentation of his project in the foreword of *Begriffsschrift* and emphasizes the link he draws between conceptual content and inference, he is deliberately leaving out Frege's self-confessed motivation for embarking on this theoretical enterprise. Let us quote directly from Frege:

> "Damit sich hierbei nicht unbemerkt etwas Anschauliches eindrängen könnte, musste Alles auf die Lückenlosigkeit der Schlusskette ankommen." (Frege 1879, iv)

In *Grundlagen* Frege openly declares that this inquiry about possible gaps in the inferential chain must be *logical*:

> "Soviel wird man zugeben, dass jede Untersuchung über die Bündigkeit einer Beweisführung [...] logisch sein muss." (Frege 1891, *Einleitung*)

And speaking about the purport of the *Begriffsschrift*:

> "Sie soll also zunächst dazu dienen, die Bündigkeit einer Schlusskette auf die sicherste Weise zu prüfen." (Frege 1879, iv)

So the driving idea of the *Begriffsschrift* is to provide a notation helpful in preventing gaps in inferential chains, which arise if empirical information sneaks in into the proof. This is required if arithmetic is to be completely reduced to logical principles.

Now it would be extremely difficult to reconcile these remarks with a material notion of inference and, unfortunately, Brandom does not even attempt it. The meaning of non-logical vocabulary essentially involves contingent information that validates the steps in material reasoning, whereas Frege strives to reduce arithmetic to "laws of thought that go beyond all

particularity" (ibid.). There is no natural way for a material inference to be "gapless". However, it is in this very context that Frege characterizes his notion of conceptual content in terms of inferential import. So Brandom cannot be right when he claims that the notion of inference involved in this definition is not that of logical inference.

There may be one last resort for those who defend the presence of some kind of material inference in Frege: they might argue that under *logical inference* Frege also understands the necessary inferences validated by the content of the non-logical concepts involved in them, such as the inference from "x is red" to "x is coloured". However, I doubt that this way out is viable. First, I know of no passage in Frege that could bear this reading out expressly. On the contrary, when he speaks about logic, he uses metaphors such as "pure thinking", "avoidance of intuitivity", "beyond all particularity" etc., which are difficult to reconcile with this reading. Second, Frege states in *Grundlagen* that nothing follows from particular facts in the absence of general covering law (Frege 1884, §3). And third, in the same breath, Frege separates general laws pertaining to special sciences from those that are genuinely logical. The inference mentioned above may be construed either as vindicated by colour theory or as based on a definition. In neither case could it be regarded by Frege as logical.

The weak reading of Frege's definition of conceptual content appears as a natural candidate to accommodate these results, as it avoids all the interpretive problems we have mentioned so far and does not give rise to any new ones.

4. Expressivism about logical vocabulary

I have already mentioned that

(3) Logical vocabulary makes explicit the implicit material inferential roles of non-logical expressions.

crucially depends on (1) and (2), so if Frege did not hold these two in the first place, there is no way he could possibly give his support to any such claim. To understand why the arguments that Brandom offers for attributing (3) to Frege are incorrect, we first have to acquire a deeper grasp of what (3) amounts to in the context of Brandom's system. The key to its meaning resides in the notion of *expression.*

Brandom himself uses the term 'expression' in two different ways – let us call them 'expression_1' and 'expression_2'. In the first, non-specialized sense, sentences express_1 the

conceptual content associated with them (Brandom 1994, 91). Within Brandom's inferentialist framework inspired by Dummett, this conceptual content amounts to validating the inference from the circumstances to the consequences of a typical utterance of the sentence.

However, it is the other meaning of 'expression' that is relevant for interpreting (3). In this sense, logical vocabulary allows us to express$_2$ or *make explicit* the conceptual content of (i.e. the inferences validated by) non-logical concepts (Brandom 1994, 108). For example, a conditional A• • • • • • $_1$ the conceptual content that validates the transition from A to B, but it also expresses$_2$ the conceptual contents of A and B themselves, by stating, respectively, some of their consequences and circumstances of application.

Since Brandom himself uses this notion in two different ways, he should be extremely careful when interpreting what Frege means by expression (*Ausdruck*). Nevertheless, the opposite is true. Brandom seems to take for granted that whenever Frege speaks about expression, he has expression$_2$ in mind. He provides barely any argument to this effect.

This premise in turn leads to one of Brandom's most outlandish claims about Frege. He suggests that according to Frege it is only in the holistic context of the whole *Begriffsschrift* that the content of logical vocabulary is specified (Brandom 1994, 96 & 114). But this is plainly wrong: Frege specifies the content of each piece of logical vocabulary he introduces by explicit definitions in terms of affirmation (*Bejahung*) conditions. Frege's proto-soundness proofs for his system make no sense under Brandom's interpretation.

Just how wide of the mark this goes transpires by what Brandom has to say about Frege's criticism of definitions that are not used in actual systematic developments of a theory. Brandom claims that the reason for Frege's disapproval of such definitions is that "their content is left entirely indeterminate" (Brandom 1994, 96). The remarks in *Grundlagen*, however, are directed against the encroachment of psychologism into the realm of mathematics, and never does Frege affirm that the content of those would-be definitions is left unspecified. On the contrary, he admits they may be germane for different purposes (Frege 1884, *Einleitung*).

If we interpret Frege's 'expression' as 'expression$_1$', we encounter no such problems. Moreover, Frege himself claims that the difference between natural language and his logical notation "only" involves a degree of precision in expressing conceptual contents. At no point is there any talk of semantic ascent, as Brandom would have it.

5. Conclusion

I believe to have made a strong case against Brandom's construal of the project of

Begriffsschrift. If an inferentialist interpretation of young Frege's philosophy is possible at all, it can draw no support from Brandom's claims. They seem to be a product of wishful thinking rather than serious analysis. It is our duty as philosophers to denounce such malpractice.

This paper makes no claim about the theoretical feasibility of inferentialism itself, which remains an interesting theoretical alternative. However, its credentials cannot be enhanced by spurious attributions to great philosophers of the past.

REFERENCES

Beaney, M. 1996, *Frege - Making Sense*, London, Duckworth.
Brandom, R. 1994, *Making it Explicit*, London, HUP.
Brandom, R. 2001, *Articulating Reasons*, London, HUP.
Dummett, M. 1981, *Frege. Philosophy of Language*, London, Duckworth.
Dummett, M. 1991, *Frege. Philosophy of Mathematics*, London, Duckworth.
Frege, G. 1879, *Begriffsschrift*, Halle a/S., Verlag von Louis Nebert.
Frege, G. 1884, *Grundlagen der Arithmetik*, Breslau.
Frege, G. 1892, 'Über Begriff und Gegenstand', *Vierteljahrsschrift für wissenschaftliche Philosophie* XVI, 192-205.

INFORMATION-THEORETIC APPROACHES TO LOGICAL RELATIONS

JOHN CORCORAN† and JOSÉ MIGUEL SAGÜILLO‡
†Department of Philosophy, University at Buffalo, Buffalo, NY 14260, USA
‡Department of Logic, University of Santiago de Compostela, Santiago 15782, Spain

ABSTRACT: This paper discusses the traditional logical relations holding or not between "corners" of squares of oppositions, between members of postulate sets, and more generally between one given proposition and another. In the last half of the 1900s, following the stunning successes of logicians such as Alfred Tarski and David Hilbert, such relations were often treated model-theoretically or proof-theoretically. We approach the subject from an information-theoretic perspective which is entirely compatible with other perspectives and which in fact complements and clarifies them. Logicians sympathetic to information-theoretic approaches include many of the giants of the last half of the 1800s: George Boole, Augustus De Morgan, Charles Peirce, Stanley Jevons, and others who worked in logic before model theory and proof theory were developed. Ideally, practice precedes and grounds theory. Ideally, theory explains and justifies practice.
KEY WORDS: information, implication, consequence, logical relation, logical form, formal, contentual.

Terminology: The following are regarded as synonymous in this paper: P implies Q, P is sufficient for Q, Q follows from P, and Q is a consequence of P. The adjective *logical* and the adverb *logically* are understood. Likewise, the following are synonymous: P does not imply Q, P is insufficient for Q, Q does not follow from P, and Q is not a consequence of P. Again, the following are synonymous: Q contains some information, Q is informative. Finally, the following are synonymous: Q contains no information, Q is uninformative, Q is tautological. For convenience, we restrict the propositions under discussion to those expressible in a given first-order language which remains fixed for the entire discussion.

Logical relations: In this paper, logical consequence and the converse relation of implication, contradiction and the complementary relation of compatibility, independence and the opposite relation of logical equivalence are all logical relations between propositions. This paper uses the expression *logical relation* as it is found used in the literature of early twentieth-century logic by colleagues of American Postulate Theorists—often in connection with relations between members of postulate sets and between members of "squares of opposition". Logical relations are thus homogeneous, two-place relations; the subject or first argument is always a proposition and the object or second argument must also be a proposition—not necessarily a different proposition. Logical relations relate propositions pairs, that is, when they hold, they relate the first member of a pair to the second.

Relations these logicians considered to be logical are all found to share two properties: "formality" and "contentuality". A *formal* relation is one that relates every pair that is in the same logical form as a pair that it relates. Formal relations "hold in virtue of form". A *contentual* relation is one that relates any pair whose members are respectively logically equivalent to the corresponding members of a pair that it relates. Contentual relations "hold in virtue of content". The relation known as formal implication—also known as logical implication—is both formal and contentual.

Relations these logicians considered non-logical in the relevant sense are found to lack one or the other of these two properties. The relation that relates a disjunction to either of its disjuncts is non-logical; it is formal but not contentual. This is the relation that relates X to Y if and only if X is a disjunction having Y as one of its disjuncts. For example, P and Q are both disjuncts of (P V Q); but their respective logical equivalents ~~P and ~~Q are not disjuncts of (P V Q). Material implication is non-logical; it is contentual but it is not formal. It is contentual because every proposition logically equivalent to a proposition materially implying a given proposition materially implies any proposition logically equivalent to the given proposition. The fact that material implication is not formal is well known, but a counterexample is in order anyway. "1 < 0"—which is false—is in the same form as "0 < 1"—which is true; "2 < 3"—which is true—is in the same form as "3 < 2"—which is false. Moreover, the pair ("1 < 0", "2 < 3") is in the same logical form as the pair ("0 < 1", "3 < 2"). But "1 < 0" materially implies "2 < 3", while "0 < 1" does not imply "3 < 2".

Some non-logical relations relevant to logical considerations are neither formal nor contentual; e.g. the relation of an axiom to a theorem of a particular axiomatic theory. In the sense of this paper—following Corcoran 2001—in order for a relation (from a proposition to a proposition) to be *logical* (in the relevant sense) it is necessary and sufficient that it be both formal and contentual.

Logical form: Let *L* be an arbitrary first-order language. Any one-one function from the vocabulary (set of characters) of *L* onto itself is *category-preserving*, *categorial*, or *a categorial*, if it carries every non-logical constant to another of the same syntactic category and carries every other character to itself. Any categorial *f* extends naturally and uniquely to the set of sentences: if *s* is a sentence of *L*, then the *n*-th character-occurrence in *fs* is *fk* where *k* is the character occurring in the *n*-th place in *s*. Any [extended] categorial extends naturally and uniquely to sentence pairs (*p*, *q*), to sets *P* of sentences, to [premise-conclusion] arguments (*P*, *c*) where *P* is a set [premise-set] and *c* a sentence [conclusion], and to "discourses" which include formal deductions whether "axiomatic" or "natural". A given sentence *p* is said to be *formally similar to* (or *formally identical to* or to have the *same logical form as*) a sentence *q* if there is a categorial *f* such that *fp* is *q* itself or an alphabetic variant of *q*. Formal similarity extends to pairs, sets, arguments, discourses, etc. Note that for two pairs to be in the same logical form it is necessary but not sufficient for the respective members to be in the same logical form. The respective members of ("0 = 1", "1 = 0") are in the same logical

form as those of ("0 = 1", "1 = 2")—but the pairs do not have the same form. For a discussion of alphabetic variants, see Quine 1940, §21.

Two propositions *are in the same logical form* if first-order sentences expressing them are in the same logical form. Two proposition pairs *are in the same logical form* if pairs of first-order sentences expressing them are in the same logical form. For more on logical form, see Corcoran 1989, 27–28; Corcoran 1994, 37–39; Corcoran 2004, 445; and the entry "logical form" in the 1999 *Cambridge Dictionary of Philosophy.* For more on the sentence-proposition distinction, see Corcoran 1989 and Corcoran 2009.

Among the several obvious metatheorems are analogues of all of the traditional principles of form: every two propositions having the same form are both tautological or both informative (non-tautological); every two sets having the same form are both consistent or both inconsistente; every two arguments in the same form are both valid or both invalid; and every two discourses in the same form are both formal deductions (in a given system) or neither is. This framework enables restatement of the Hilbert-Bernays-Quine theorem that, in a sufficiently rich interpreted language without identity, in order for a propostion to be consistent (respectively, tautological) it is necessary and sufficient for it to be in the same logical form as some true proposition (respectively, no false proposition).

Implication: According to information-theoretic approaches, P implies Q iff P contains all information in Q. The condition that P contains all information in Q divides into two:

1) P contains all information in Q and more.

2) P contains all information in Q and no more.

These two mutually exclusive cases have both been studied and appropriately named; the first is *superimplication*, the second [logical] *equivalence*.

Superimplication: According to information-theoretic approaches, P superimplies Q iff P contains all information in Q and more—in other words, iff P contains all information in Q but Q does not contain all information in P—in still other words, iff P implies Q but Q does not imply P.

Logical equivalence: According to information-theoretic approaches, P is equivalent to Q iff P contains all information in Q and no more—in other words, iff P contains exactly the same information as Q—in still other words, iff P implies Q and conversely Q implies P.

P implies Q iff either P superimplies Q or P is equivalent to Q.

Notice that if Q contains no information, every proposition P implies Q. This means that if some proposition P does not imply Q, Q is informative. In order not to follow, a proposition must be informative.

Insufficiency: The condition that P is insufficient for Q, that P does not contain all information in Q, that is, Q contains information not in P, also divides into two:

1) Q is informative and P contains some but not all information in Q.

2) Q is informative and P contains none of the information in Q.

These two cases have not been studied extensively and appropriate terminology does not exist; the first is a species of what Corcoran called [informational] connectedness; the second is a species of what Corcoran called [informational] unconnectedness.

Recall that if P is insufficient for Q, that is, if P does not imply Q, then Q is informative.

Historically, if P is insufficient for Q, then it was misleadingly said that Q is independent of P. For example, this terminology would require saying that many conjunctions are independent of their conjuncts and that every false proposition is independent of its negation. It is not to the point to discuss this historical oddity here.

Connectedness: According to information-theoretic approaches, P is connected to Q iff P contains some information in Q. As usual in logic, 'some' is used in the broad sense "at least some", that is, "some or some and all". Thus, P is connected to Q iff P contains at least some of the information in Q.

Thus, P is connected to Q iff either P contains some *and all* information in Q or P contains some *but not all* information in Q. In other words, P is connected to Q iff either P implies Q and contains some of Q's information or P does not imply Q *but contains* some of Q's information.

Unconnectedness: P is unconnected to Q iff P contains no information in Q. Thus, P is unconnected to Q iff either P implies Q and Q contains no information or P does not implies Q and Q contains some information but none in P.

The literature amply demonstrates the heuristic effectiveness of information-theoretic approaches. See for example Sagüillo 1997, 2000, or 2009.

Acknowledgements: Many thanks to Frango Nabrasa, Joaquin Miller, Roberto Torretti, and others. This paper is framed under research projects FFI2008-00085, y FFI2009-08828 of Ministerio de Ciencia e Innovación.

REFERENCES

Audi, R, (Editor) 1999, *The Cambridge Dictionary of Philosophy*, Cambridge, Cambridge University Press, second edition.

Corcoran, J. 1989, Argumentations and Logic, *Argumentation* 3, 17–43.

Corcoran, J. 1994, Argumentaciones y lógica, *Agora* 13/1, 27–55. Translation by R. Fernandez and J. Sagüillo of a revised and expanded version of Corcoran 1989.

Corcoran, J. 1995, Information Recovery Problems, *Theoria* 10, 55–78.

Corcoran, J. 1998, Information-theoretic logic, C. Martínez, U. Rivas and L. Villegas-Forero, eds., *Truth in Perspective*, Aldershot, UK: Ashgate, 113–135.

Corcoran, J. 1999, Information-theoretic logic and transformation-theoretic logic, *Fragments in Science*, edited by M. Ram, World Scientific Publishing Company, River Edge, New Jersey, 25–35.

Corcoran, J. 2001, What are logical relations? *Bulletin of Symbolic Logic* 7, 105–106.

Corcoran, J. 2004, First-order Logical Form. *Bulletin of Symbolic Logic* 10, 445.

Corcoran, J. 2007, Information-theoretic properties of truth-functional connectives. *Bulletin of Symbolic Logic.* 13, 405.

Corcoran, J. 2008, An information-theoretic approach to first-order logic. *Bulletin of Symbolic Logic.* 14 (2008) 415–6.

Corcoran, J. 2009, Sentence, Proposition, Judgment, Statement, and Fact: Speaking about the Written English Used in Logic. *The Many Sides of Logic*. Editors: W. A. Carnielli, M. E. Coniglio, I. M. Loffredo D' Ottaviano. Series "Studies in Logic". London: College Publications, 71–103.

Quine, W. V. 1940, *Mathematical Logic*. Cambridge, MA: Harvard UP.

Sagüillo, J. M. 1997, Logical consequence revisited. *The Bulletin of Symbolic Logic* 3, 216-241.

Sagüillo, J. M. 2000, Domains of sciences, universes of discourse and omega arguments. *History and Philosophy of Logic* 20, 267-290.

Sagüillo, J. M. 2009, Methodological practice and complementary concepts of logical consequence: Tarski's Model-Theoretic Consequence and Corcoran's Information-Theoretic Consequence. *History and Philosophy of Logic* 30, 21–48.

Scanlan, M. 1991, Who were the American Postulate Theorists? *Journal of Symbolic Logic* 56, 981-1002.

LA HISTORIA DE LA LÓGICA COMO LA HISTORIA DE LOS ARGUMENTOS ONTOLÓGICOS

MANUEL CRESCENCIO MORENO
U. de Salamanca

ABSTRACT: The ontological arguments are arguments that pretend to proof the existence of God with the use of reason alone. They are considered *a priori* and for that matter are concerned with logic, if we consider logic is also *a priori*. In this paper I want to show the parallelisms between the formulation of the ontological arguments and the development of logic by focusing on their mutual influences. In the wider section of the article I try to make use of different logic systems for formalising the variety of ontological arguments, for which I have searched a valid proof. In spite of the validity of the arguments, it must be said that this does not imply their cogency because the premises are not entirely justified and we can even find the fallacy known as *petitio principii*. In any case, I would like to show that the formalization of the ontological arguments enables us to deal with many topics in logic such as first order logic, higher order logic and modal logic among others.

Key words: Argumento ontológico, Dios, Lógica de primer orden, lógica de segundo orden, lógica modal.

1. Introducción

Resulta bastante común encontrarnos con determinados ejemplos que nos identifican con rapidez el sistema lógico en el que nos movemos. Cuando escuchamos hablar de "Todos lo hombres son mortales...", inmediatamente nos vamos a la silogística, si se nos comienza a decir que "El Rey de Francia es calvo", nos trasladamos al tratamiento de las descripciones definidas en la lógica de primer orden, y si se nos hace referencia a que "el número de los planetas es necesariamente impar", todos nos situamos inmediatamente en un ambiente modal. Cada ejemplo es representativo de un tipo de lógica concreto y, podría decirse que, si existiese una heráldica para la lógica, el emblema de muchos sistemas lógicos estaría ya definido por los diferentes ejemplos utilizados.

Pero, ¿existe algún ejemplo que nos pueda servir para todas las lógicas? La finalidad del presente texto es dar una respuesta afirmativa a esta pregunta, para tratar de identificar en la argumentación ontológica la fuente de numerosos ejemplos que han recorrido la historia del pensamiento occidental y que, en la actualidad, pueden ser formalizados utilizando diversos sistemas formales. Con la aparición de algunas nuevas lógicas, los argumentos ontológicos

vuelven a renacer y, si bien su valor de convicción puede ser obviado, sin embargo, su riqueza heurística, es algo que debe ser tomado en consideración. En este sentido, pretendo desarrollar a continuación dos líneas que señalen el uso fructífero de los mismos:

1. La primera tratará de señalar la influencia mutua entre argumentos ontológicos y sistemas lógico-formales.

2. La segunda estudiará la "cantidad" de lógica necesaria para formalizar los argumentos más significativos, de modo que cada argumento sea representativo de una lógica determinada.

2. Influencia mutua lógicas-ontológicos

En este primer punto desarrollo en un primer apartado cómo la creación de nuevas lógicas ha propiciado tanto la adaptación de antiguas versiones de pruebas *a priori* para demostrar la existencia de Dios como la creación de alguna nueva y, a continuación, expongo cómo la búsqueda de una prueba válida ha contribuido a enriquecer determinados sistemas lógicos.

2.1. Influencia de la lógica sobre los argumentos ontológicos

Con la aparición de la lógica modal, algunos autores pretenden recuperar la conocida argumentación que San Anselmo presenta en el capítulo II de su *Proslogion* (Anselmo, 1952) desde un punto de vista formal, aunque inspirándose principalmente en otro argumento anselmiano que identifican en el capítulo III de dicha obra y que está más en consonancia con un tratamiento modal de la prueba.

En este sentido, la propuesta de Hartshorne (Hartshorne, 1962) presenta la siguiente traducción modal de los axiomas de Anselmo (donde se ha modificado la definición de Dios como *el ser mayor que el cual nada cabe pensar —id quo maius cogitari nequit—* por *el ser perfecto*):

- El axioma que señala que "un ser perfecto existe en la mente", se convierte en "un ser perfecto es posible".

- La suposición que afirma que "no puede darse el caso que un ser perfecto exista en la mente y que no exista en la realidad" se convierte en "un ser perfecto debe existir necesariamente".

Si introducimos en nuestro lenguaje el predicado Px para expresar que x *es un ser perfecto*, nos encontramos con la siguiente formalización:

Un ser perfecto existe en la mente, queda como

$$\Diamond \exists x Px \tag{1}$$

y, *un ser perfecto debe existir necesariamente*, se expresa como

$$\Box(\exists x Px \rightarrow \Box \exists x Px) \tag{2}$$

Con lo anterior es posible deducir que *hay al menos un ser que es perfecto*: $\exists x Px$

$$\{\Diamond \exists x Px, \Box(\exists x Px \rightarrow \Box \exists x Px)\} \vdash \exists x Px \tag{3}$$

En esta misma línea de una reinterpretación modal de la prueba de San Anselmo pueden situarse además los trabajos de Malcolm, aunque tratada en un lenguaje informal (Malcolm, 1960), y de Plantinga (Plantinga, 1974), en los que no nos detenemos en este momento.

Además de la reformulación de la prueba clásica del *Proslogion*, en el siglo XX también nos encontramos con la creación de alguna nueva demostración: es el caso de la prueba de Kurt Gödel (Gödel, 1995, 403–404) que, aunque inspirada en Leibniz —porque sigue el esquema propuesto por éste de que hay que llegar a mostrar que *la existencia de Dios es posible* puesto que *si es posible que Dios existe, entonces es necesario que exista*—, sin embargo, contiene varios elementos originales que hacen de ella una prueba novedosa. Entre estos, el más característico es que está formulada totalmente en un lenguaje formal —a diferencia del resto de argumentos clásicos, que parten de una exposición en el lenguaje natural— y, además, contiene varios supuestos que tratan de explicar cómo funcionan las *propiedades positivas* y cómo se definen las propiedades de *esencia* y *existencia necesaria*.

2.2. Influencia de los argumentos ontológicos sobre la lógica

Sin la pretensión de ser exhaustivo, es posible identificar algunos aspectos en los que, la búsqueda de una prueba válida, ha conducido a la realización de ciertas aportaciones significativas dentro del campo de la lógica.

Una de ellas es la enunciación de alguna proposición que será considerada posteriormente un teorema lógico. Este es el caso de la afirmación de Leibniz (Gerhardt, 1978, vol. I, 188) que reza: *si es posible que el ser necesario existe, entonces el ser necesario existe*, la cual puede

formalizarse en un teorema del sistema S5 de lógica modal:

$$\Diamond\Box p \rightarrow \Box p \tag{4}$$

Asimismo, aunque en un contexto *a posteriori*, puesto que lo que se pretende es compatibilizar la existencia de Dios con la presencia del mal en el mundo, Leibniz expone su teoría de los mundos posibles (Leibniz, 1985), adelantándose a la semántica vigente de las lógicas modales.

Otra de las aportaciones concierne al tratamiento de la *existencia* como predicado. La dura crítica recibida por Kant y Frege respecto a dicha asunción en los argumentos ontológicos no ha sido definitiva, puesto que se ha impuesto como un instrumento útil en el desarrollo de la semántica posibilista en las lógicas modales y en la formación de las lógicas libres. De este modo, se permite diferenciar entre *denotación* y *existencia*, lo cual no es posible con la sola presencia del cuantificador existencial $\exists$ y que, sin embargo, está tan presente en nuestro lenguaje cotidiano.

3. Lógicas distintas para argumentos distintos

Para tratar este punto, más amplio, me detengo en la traducción de los argumentos a un lenguaje formal lo más adecuado posible al texto original. Aun pretendiendo ser lo más próximo posible a la literalidad del texto, he de decir que, en algunos casos, he debido introducir en la formalización algunos axiomas añadidos que no se encuentran explícitamente en el texto de los autores, debido a que hay algunas premisas que éstos han asumido implícitamente en las demostraciones.

A pesar de esto, creo que los argumentos formalizados no desvirtúan el sentido ni de los originales, ni de la filosofía del autor. De todos modos, no entraré en un análisis crítico sobre la aceptación de las premisas, ni sobre la filosofía de fondo, ya que esto excede el planteamiento que propongo en este artículo.

En el proceso de realizar las traducciones formales de los diversos argumentos, nos será posible ir reconociendo en ellas el funcionamiento de distintas lógicas, que van desde la lógica de primer orden hasta una lógica modal cuantificada de orden superior, por lo que quedará puesto de manifiesto cómo las diversas elaboraciones de una prueba *a priori* para demostrar la existencia de Dios, nos permiten acercarnos al estudio de una gran diversidad de sistemas lógicos; pudiendo decirse, incluso, que es posible un estudio de la lógica y de su desarrollo a

través del mero análisis formal de los argumentos ontológicos.

3.1. La lógica de primer orden y el argumento de San Anselmo

La demostración de Anselmo, considerada la más clásica de todas las pruebas, requiere, para una formalización adecuada, de una lógica de primer orden. De este modo, utilizando el siguiente lenguaje formal,

$$Mx := x \text{ existe en la mente, o bien, } x \text{ puede pensarse}$$

$$Rx := x \text{ existe en la realidad}$$

$$Sxy := x \text{ es superior a } y$$

$$Dx := x \text{ es divino}$$

podemos formalizar el argumento de la siguiente manera:

Premisas:

Definición $\alpha :=$ Dios, o mejor, lo divino, es aquello mayor que lo cual nada cabe pensar, es decir, es aquello que puede pensarse y que es superior a cualquier otra cosa que puede pensarse.

$$\forall x(Dx \leftrightarrow Mx \wedge \neg\exists y(My \wedge Syx)) \tag{5}$$

Axioma $\beta :=$ Algún ser existe en la mente y en la realidad.

$$\exists x(Mx \wedge Rx) \tag{6}$$

Axioma $\gamma :=$ Todo ser que pueda ser pensado y exista en la realidad es superior a cualquier otro ser que pueda ser pensado y no exista en la realidad.

$$\forall x(Mx \wedge Rx \rightarrow \forall y(My \wedge \neg Ry \rightarrow Sxy)) \tag{7}$$

Axioma $\delta :=$ Dios puede ser pensado.

$$\exists x(Dx \wedge Mx) \tag{8}$$

Conclusión. $\varphi :=$ Dios existe en la mente y en la realidad.

$$\exists x(Dx \wedge Mx \wedge Rx) \tag{9}$$

Dado lo anterior, tenemos que se cumple la siguiente relación de deducibilidad:

$$\{\alpha, \beta, \gamma, \delta\} \vdash \varphi \tag{10}$$

3.2. La lógica de segundo orden y el argumento de Descartes

El argumento de Descartes (Descartes, 1977, 77-80) trata la *perfección* como una propiedad de propiedades, lo cual nos lleva a movernos dentro de un lenguaje de segundo orden que requiere de:

Dos predicados monarios de primer orden:

$$Hx := x \text{ es un ser supremo}$$

$$Ex := x \text{ posee existencia}$$

Un predicado de segundo orden para formalizar la perfección:

$$\mathcal{F}X := X \text{ es una perfección}$$

Con lo anterior, podemos construir formalmente el argumento de Descartes:

Premisas:

Definición $\alpha :=$ Un ser supremo es un ser que posee todas las perfecciones.

$$\forall x(Hx \to \forall X(\mathcal{F}X \to Xx)) \tag{11}$$

Axioma $\beta :=$ La existencia es una perfección.

$$\mathcal{F}E \tag{12}$$

Conclusión. $\varphi :=$ Un ser supremo tiene que existir.

$$\forall x(Sx \to Ex) \tag{13}$$

En realidad, este argumento no acaba probando nada concreto debido a que puede ser vacuamente verdadero.

3.3. Leibniz y el giro modal del argumento

La principal aportación de Leibniz es la consideración de que, si partimos de la posibilidad de Dios, podemos llegar a la necesidad de la existencia de Dios. Para ello, lo único que debemos hacer es añadir como axioma la proposición que afirma que Dios no es un ser contingente. Si introducimos en nuestro lenguaje el símbolo p para expresar que *Dios existe*, podemos construir el siguiente argumento válido en **S5**:

Premisas:

Axioma $\alpha :=$ Es posible que Dios exista.

$$\Diamond p \tag{14}$$

Axioma $\beta :=$ Dios no es un ser contingente, es decir, no se da el caso que la existencia de Dios sea posible y que no sea necesaria.

$$\neg(\Diamond p \wedge \neg\Box p) \tag{15}$$

Lo que en realidad nos está diciendo este axioma es que la existencia de Dios es imposible o es necesaria. De hecho, β es equivalente a:

$$\neg\Diamond p \vee \Box p \tag{16}$$

Conclusión. $\varphi :=$ Dios existe.

$$p \tag{17}$$

Entonces

$$\{\Diamond p, \neg(\Diamond p \wedge \neg\Box p)\} \vdash p \tag{18}$$

Esto nos muestra la importancia que tiene la asunción de admitir la posibilidad de la existencia de Dios ya que, si por el contrario, partimos de que es posible que Dios no existe llegamos a la conclusión de que Dios no existe.

$$\{\Diamond\neg p, \neg(\Diamond p \wedge \neg\Box p)\} \vdash \neg p \tag{19}$$

3.4. La lógica modal y el argumento de Hartshorne

Aunque la versión de Hartshorne del argumento de Anselmo que hemos presentado antes contenía cuantificadores, sin embargo, es posible formularlo adecuadamente en una lógica modal

proposicional en la que la fórmula $\exists x Px$ quede abreviada por la letra p, que puede traducirse por: *un ser perfecto existe*. El argumento anterior se nos mostraría entonces del siguiente modo:

Premisas:

$\alpha :=$ Un ser perfecto existe en la mente.

$$\Diamond p \tag{20}$$

$\beta :=$ Un ser perfecto debe existir necesariamente.

$$\Box(p \to \Box p) \tag{21}$$

Conclusión. $\varphi :=$ Un ser perfecto existe.

$$p \tag{22}$$

No obstante, también puede realizarse una lectura de las premisas de Hartshorne, que requiere de una lógica modal de primer orden para su formalización (Sobel, 2004, 86-87). En este caso, la premisa β, admitiría una interpretación según la cual: *necesariamente, un ser perfecto debe necesariamente existir y ser perfecto*. El argumento se presenta, entonces, con cuantificadores y operadores modales y con los predicados P para *es un ser perfecto* y E para la *existencia*:

Premisas:

$\alpha' :=$ Un ser perfecto existe en la mente.

$$\Diamond \exists x Px \tag{23}$$

$\beta' :=$ Un ser perfecto debe existir necesariamente.

$$\Box \forall x(Px \to \Box(Ex \wedge Px)) \tag{24}$$

Conclusión. $\varphi' :=$ Un ser perfecto existe.

$$\exists x Px \tag{25}$$

3.5. La lógica modal de orden superior y el argumento de Gödel

En el argumento de Gödel se utiliza la cuantificación sobre propiedades, lo cual nos lleva a afirmar que la lógica utilizada es de segundo orden. Además, se hace uso de operadores modales. Estaremos utilizando, por tanto, un lenguaje modal cuantificado de segundo orden. Dicho lenguaje contiene los siguientes predicados de primer orden y de orden superior:

$$Dx := x \text{ es un ser divino}$$

$$X \text{ Ess } x := X \text{ es la esencia de } x$$

$$Nx := x \text{ tiene existencia necesaria}$$

$$\mathcal{P}X := X \text{ es una propiedad positiva}$$

Premisas:

Definición $\alpha :=$ Divino es aquello que posee todas las propiedades positivas.

$$\forall x[Dx \leftrightarrow \forall X[\mathcal{P}X \rightarrow Xx]] \tag{26}$$

Definición $\beta :=$ La esencia de un objeto es aquella propiedad que implica todas las propiedades que dicho objeto posee.

$$\forall X \forall x[X \text{ Ess } x \leftrightarrow Xx \wedge \forall Y(Yx \rightarrow \Box \forall y[Xy \rightarrow Yy])] \tag{27}$$

Definición $\gamma :=$ Un objeto posee la propiedad de la existencia necesaria si su esencia es necesariamente ejemplificada.

$$\forall x[Nx \leftrightarrow \forall X[X \text{ Ess } x \rightarrow \Box \exists x Xx]] \tag{28}$$

Axioma $\delta :=$ O bien una determinada propiedad es positiva o bien lo es su contraria.

$$\forall X(\mathcal{P}(\neg X) \leftrightarrow \neg \mathcal{P}(X)) \tag{29}$$

Axioma $\varepsilon :=$ Toda propiedad que es implicada necesariamente por una propiedad positiva también es positiva.

$$\forall X \forall Y[\mathcal{P}X \wedge \Box \forall x[Xx \rightarrow Yx] \rightarrow \mathcal{P}Y] \tag{30}$$

Axioma $\zeta :=$ La conjunción de cualquier número de propiedades positivas también es positiva.

$$\forall X \forall Y [\mathcal{P}X \wedge \mathcal{P}Y \rightarrow \mathcal{P}(X \wedge Y)] \tag{31}$$

Axioma $\eta :=$ Toda propiedad positiva es necesariamente positiva.

$$\forall X [\mathcal{P}X \rightarrow \Box \mathcal{P}X] \tag{32}$$

Axioma $\theta :=$ La existencia necesaria es una propiedad positiva.

$$\mathcal{P}N \tag{33}$$

Conclusión. $\varphi :=$ Es necesario que exista al menos un ser que sea divino.

$$\Box \exists x D(x) \tag{34}$$

En el sistema **S5**, tenemos que

$$\{\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta, \theta\} \vdash \varphi \tag{35}$$

Dicho argumento también puede formalizarse en una lógica intensional que contenga términos intensionales y el operador λ, como Fitting ha mostrado en su obra (Fitting, 2002).

4. Conclusión

La exposición anterior nos permite concluir con la idea de que se han ido utilizando, con el paso del tiempo, lógicas cada vez más expresivas para formalizar los argumentos ontológicos, lo cual los convierte en merecedores de un gran interés, dado que, seguramente, no existe otro argumento que haya sido tan ampliamente analizado en épocas tan distintas y en sistemas lógicos tan diversos como este.

REFERENCIAS

Anselmo, S. (1952). *Obras completas de San Anselmo. Volumen I.* Madrid: BAC.

Descartes, R. (1977). *Discurso del Método. Meditaciones Metafísicas. Reglas para la dirección del espíritu. Principios de la Filosofía.* México: Porrúa.

Fitting, M. (2002). *Types, Tableaus, and Gödel's God.* Dordrecht: Kluwer Academic Publishers.

Gerhardt, C. I. (Ed.). (1978). *Die philosophischen Schriften von G. W. Leibniz.* Hidesheim: Georg Olms.

Gödel, K. (1995). *Collected Works. Volume III: Unpublished Essays and Lectures.* New York: Oxford University Press.

Hartshorne, C. (1962). *The Logic of Perfection.* La Salle, IL: Open Court.

Leibniz, G. W. (1985). *Theodicy: Essays on the Goodness of God, the Freedom on Man and the Origin of Evil.* La Salle, IL: Open Court.

Malcolm, N. (1960). Anselm's ontological arguments. *The Philosophical Review*, *69*, 41–62.

Plantinga, A. (1974). *The Nature of Necessity.* Oxford: Oxford University Press.

Sobel, J. H. (2004). *Logic and Theism.* New York: Cambridge University Press.

LA *JERARQUÍA DE GRZEGOPRZCYK* EN TEORÍA DE CATEGORÍAS

JOAQUÍN DÍAZ BOÏLS
U. de Valencia

ABSTRACT: A new categorical setting is defined in order to characterize the subrecursive classes belonging to the *Grzegorzcyk Hierarchy*. This is achieved by means of *coercion functors* over a symmetric monoidal category endowed with certain recursion schemes that imitate the bounded recursion scheme. This is donde to get a categorical counterpart of generalized *safe composition* and *safe recursion*.[1]

Key words: Symmetric monoidal category, Lawvere Theory, Safe composition, Safe recursion.

Distintas clases de funciones recursivas han sido caracterizadas por medio de Teoría de Categorías. Ello se ha logrado considerando categorías cartesianas a las que se ha añadido alguna estructura más y dotadas de un esquema de recursión. Es sabido que puede obtenerse otras clases *por encima de* (conteniendo a) $\mathscr{PR}$ añadiendo alguna estructura más a la categoría sobre la que se trabaja: considerando por ejemplo un topos incial (en (Lambek & Scott., 1986)), una categoría cartesiana cerrada (en (Thibault, 1977)) o una categoría con límites finitos (en (Roman, 1990)).

Menos trabajo, sin embargo, se ha presentado sobre la caracterización de las clases de funciones *subrecursivas*, esto es, aquéllas contenidas en $\mathscr{PR}$ (consultar por ejemplo (Cockett & Redmond, 2010)). La razón principal es que no se dispone de un diagrama recursivo con suficiente expresividad que caracterice la recursión acotada de la que están dotadas dichas clases y que tiene el aspecto siguiente

$$\begin{cases} f(u,0) = g(u) \\ f(u,x+1) = h(u,x,f(u,x)) \\ f(u,x) \leq j(u,x) \end{cases}$$

Se hace necesaria una aproximación diferente al problema de caracterizar las clases de funciones subrecursivas, esto es, encontrar un modo categorial de expresar la semántica de caracterizaciones

[1] La presente comunicación contiene los resultados principales de la tesis dirigida por José Pedro Úbeda Rives.

tales como la *recursión ramificada* (ver (Leivant, 1994)) o la *recursión segura* (ver (Bellantoni & Cook, 1992)).

El trabajo hecho en la tesis presentada por J. R. Otto en 1995 ((Otto, 1995)) hace uso de las categorías de números ordinales para definir unas *categorías flecha* extendidas que dan lugar a *endofuntores de coerción* con idea de recoger las condiciones de ramificación de las funciones a definir, y trabajar con las estrcuturas definidas tal como se hacía en los textos primeros ((Burroni, 1986; Roman, 1989; Lambek & Scott., 1986)). Por este método se trata de caracterizar en (Otto, 1995) las clases de funciones recursivas de tiempo lineal, espacio lineal, espacio polinomial, tiempo polinomial y los niveles 2 y 3 de la *Jerarquía de Grzegorzcyk*.

Se verá, sin embargo, que las construcciones sintácticas de (Otto, 1995) con profundidad de recursión igual a 3 se reducen a los niveles anteriores y se dará una solución a tal problema.

La presente comunicación corrige los problemas señalados en el párrafo anterior, ofrece una caracterización de toda la *Jerarquía de Grzegorzcyk* en términos de Teoría de Categorías y, al mismo tiempo, ciertas herramientas que pueden revelarse útiles para posteriores caracterizaciones de más clases de funciones subrecursivas. Esto es, una aportación a una lectura más amplia de la Teoría de la Computabilidad en términos de la lógica categorial, usando la terminología de J. Lambek y P. Scott en (Lambek & Scott., 1986).

Definición 1. Para cada $n \in \mathbb{N}$ la categoría $\mathbf{n}$ es la categoría cuyos objetos son los números naturales menores que n:

$$n = \{0, 1, \ldots, n-1\}$$

y cuyas flechas, que corresponden al orden lineal de n determinada por el diagrama:

$$0 \longrightarrow 1 \longrightarrow \cdots \longrightarrow n-1,$$

serán denotadas por $m_{i,j}$ con $0 \leq i \leq j < n$ de tal modo que $m_{i,j} : i \longrightarrow j$ es la única flecha cuyo dominio es i y cuyo codominio es j.

Proposición 1. $\mathbf{n}$ *es una categoría simétrica monoidal.*[2]

Observación. Es fácil ver que $\mathbf{n}$ tiene n objetos y $n(n+1)/2$ flechas.

El monoide M_n^{op} de endomorfismos de $\mathbf{n}$ en el cual el producto fg corresponde a la composición $g \circ f$ es exactamente el conjunto de funciones monótonas de n a n. Este conjunto es, como puede observarse, el conjunto de endofuntores en $\mathbf{n}$.

[2]En adelante denominaremos *categoría SM* a una categoría simétrica monoidal.

Asimismo la 2-categoría M_n es aquella cuya celda de nivel cero es el conjunto n con las flechas $m_{i,j}$; su celda de nivel uno es el monoide de las funciones monótonas y sus flechas las transformaciones naturales entre funciones monótonas consideradas como funtores y que corresponden al orden entre las funciones monótonas inducidas por el orden lineal en n.

Vamos a establecer un conjunto de elementos de M_n^{op} a partir de los cuales pueden generarse todos los restantes por medio de la multiplicación. Este conjunto se usa en (Otto, 1995) en el caso de $n = 2$ y 3.

Definición 2. Sean para cada $0 \leq k < n-1$ los funtores $id : \mathbf{n} \longrightarrow \mathbf{n}$, $T_k : \mathbf{n} \longrightarrow \mathbf{n}$ y $G_k : \mathbf{n} \longrightarrow \mathbf{n}$ tales que para cada $j \in \mathbb{N}$:

$$id(j) = j$$

$$T_k(j) = \begin{cases} j+1 & \text{si } j = k \\ j & \text{si } j \neq k \end{cases}$$

$$G_k(j) = \begin{cases} j-1 & \text{si } j = k+1 \\ j & \text{si } j \neq k+1 \end{cases}$$

que gráficamente toman las formas

$$\begin{tikzcd}
{[T_0]} & 0 \arrow[dr] & 0 \\
 & 1 \arrow[r] & 1 \\
 & 2 \arrow[r] & 2 \\
 & \vdots & \vdots \\
 & n-1 \arrow[r] & n-1
\end{tikzcd}
\qquad
\begin{tikzcd}
{[T_1]} & 0 \arrow[r] & 0 \\
 & 1 \arrow[dr] & 1 \\
 & 2 \arrow[r] & 2 \\
 & \vdots & \vdots \\
 & n-1 \arrow[r] & n-1
\end{tikzcd}
\qquad \cdots \qquad
\begin{tikzcd}
{[T_{n-2}]} & 0 \arrow[r] & 0 \\
 & 1 \arrow[r] & 1 \\
 & \vdots & \vdots \\
 & n-2 \arrow[dr] & n-2 \\
 & n-1 \arrow[r] & n-1
\end{tikzcd}$$

y

$$
\begin{tikzcd}
[G_0] & 0 \arrow[r] & 0 & [G_1] & 0 \arrow[r] & 0 & \cdots & [G_{n-2}] & 0 \arrow[r] & 0 \\
 & 1 \arrow[ur] & 1 & & 1 \arrow[r] & 1 & & & 1 \arrow[r] & 1 \\
 & 2 \arrow[r] & 2 & & 2 \arrow[ur] & 2 & & & \vdots & \vdots \\
 & \vdots & \vdots & & \vdots & \vdots & & & n-2 \arrow[r] & n-2 \\
 & n-1 \arrow[r] & n-1 & & n-1 \arrow[r] & n-1 & & & n-1 \arrow[ur] & n-1
\end{tikzcd}
$$

En adelante nos referiremos a todos los T_k y G_k como los *funtores de coerción*.

Proposición 2. *Para cada $n \in \mathbb{N}$ el monoide M_n^{op} puede ser generado por medio del conjunto finito*

$$\{id, G_0, \cdots, G_{n-2}, T_0, \cdots, T_{n-2}\}$$

Algunas propiedades de los generadores de M_n^{op} se dan en los Teoremas siguientes.

Teorema 1. *Para todos los generadores id, T_k y G_k con $0 \leq k < n-1$ y para $m, i, i+1, j \in \mathbf{n}$*

1. *id, T_k y G_k son idempotentes para todo $0 \leq k < n-1$.*
2. *$T_k G_k = G_k$ y $G_k T_k = T_k$ cuando $0 \leq k < n-1$.*
3. *$T_{k+1} G_k = T_{k+1}$ y $G_k T_{k+1} = G_k$ cuando $0 \leq k < n-2$.*
4. *$T_k G_m = G_m T_k$ cuando $k \neq m, m+1$ y $0 \leq k < n-1$.*
5. *$T_i T_j = T_j T_i$ cuando $j \neq i-1, i+1$.*
6. *$G_i G_j = G_j G_i$ cuando $j \neq i-1, i+1$*
7. *$T_i T_{i+1} \neq T_{i+1} T_i$ y $G_i G_{i+1} \neq G_{i+1} G_i$.*
8. *$T_i T_{i+1} T_i = T_i T_{i+1}$ y $G_i G_{i+1} G_i = G_{i+1} G_i$.*

Teorema 2. *Sean j y k tal que $0 \leq j, k < n$.*

1. *Si $j \neq k$ entonces $G_k T_j G_k = G_k T_j$ y $T_k G_j T_k = T_k G_j$.*
2. *Si $k \neq j+1$ entonces $T_k G_j T_k = G_j T_k$ y $G_j T_k G_j = T_k G_j$.*

3. *Si* $k \neq j+1$ *entonces* $T_kT_jT_k = T_kT_j$ *y* $G_jG_kG_j = G_jG_k$.

A continuación definimos algunas transformaciones naturales en **n**.

Definición 3. Sean $\varepsilon_k : G_k \Rightarrow id$ y $\eta_k : id \longrightarrow T_k$ $(0 \leq k \leq n-2)$ tales que

$$\varepsilon_k(i) = \begin{cases} m_{i,i} & \text{si } i \neq k+1 \\ m_{i-1,i} & \text{si } i = k+1 \end{cases}$$

$$\eta_k(i) = \begin{cases} m_{i,i} & \text{si } i \neq k \\ m_{i,i+1} & \text{si } i = k \end{cases}$$

Teorema 3. *Toda transformación natural en* **n** *puede ser generada por medio de una composición de transformaciones naturales y por la multiplicación de una transformación natural y un funtor a partir de* T_k*,* G_k*,* ε_k*,* η_k *y las transformaciones identidad* $(k = 0, \ldots, n-2)$*.*

A continuación veremos que una $n-$Comprensión *SM*, del mismo modo que una $2-$Comprensión, es realmente un 2-funtor

$$M_n \longrightarrow \mathscr{SM}$$

donde $\mathscr{SM}$ es la 2-categoría de todas las categorías *SM*.

Para ello se necesita considerar un concepto de *modelo* que nos permita definir estructuras categoriales a partir de otras estructuras en base a unas ciertas propiedades que las nuevas *heredan* de las primeras.

El concepto de *modelo* entre 2-categorías que aquí se define puede ser expresado por medio del concepto más general de *Teoría Algebraica de Lawvere*. Los modelos son los funtores entre *teorías algebraicas de Lawvere*.

Informalmente, una categoría con una cierta estructura algebraica 2-categorial es un *modelo* de otra categoría con la misma estructura si existe un funtor entre ellas que la preserva. En particular, sea el $(4n-3)$-tuplo $(\mathbf{n}, T_k, G_k, \eta_k, \varepsilon_k)$. Diremos que, dada una categoría *SM* $\mathscr{C}$, un $(4n-3)$-tuplo $(\mathscr{C}, T_k^{\mathscr{C}}, G_k^{\mathscr{C}}, \eta_k^{\mathscr{C}}, \varepsilon_k^{\mathscr{C}})$ es un *modelo* de $(\mathbf{n}, T_k, G_k, \eta_k, \varepsilon_k)$ si $T_k^{\mathscr{C}}, G_k^{\mathscr{C}}, \eta_k^{\mathscr{C}}$ y $\varepsilon_k^{\mathscr{C}}$ satisfacen los mismos diagramas conmutativos que T_k, G_k, η_k y ε_k.

Definición 4. Una *n-Comprensión SM* es un $(4n-3)$-tuplo $(\mathscr{C}, T_k^{\mathscr{C}}, G_k^{\mathscr{C}}, \eta_k^{\mathscr{C}}, \varepsilon_k^{\mathscr{C}})$ donde[3]

1. $\mathscr{C}$ es una categoría *SM*

[3]En adelante $\top$ denotará la unidad del producto tensorial $\otimes$ en $\mathscr{C}$.

2. para cada k tal que $0 \leq k < n-1$ los funtores $T_k^{\mathscr{C}}, G_k^{\mathscr{C}} : \mathscr{C} \longrightarrow \mathscr{C}$ son *funtores SM* en el siguiente sentido:

   a) $T_k^{\mathscr{C}} \top = G_k^{\mathscr{C}} \top = \top$

   b) $T_k^{\mathscr{C}}(f \otimes Y) = T_k^{\mathscr{C}} f \otimes T_k^{\mathscr{C}} Y$ y $G_k^{\mathscr{C}}(f \otimes Y) = G_k^{\mathscr{C}} f \otimes G_k^{\mathscr{C}} Y$

   c) $T_k^{\mathscr{C}} \alpha XYZ = \alpha(T_k^{\mathscr{C}} X)(T_k^{\mathscr{C}} Y)(T_k^{\mathscr{C}} Z)$ y $G_k^{\mathscr{C}} \alpha XYZ = \alpha(G_k^{\mathscr{C}} X)(G_k^{\mathscr{C}} Y)(G_k^{\mathscr{C}} Z)$

   d) $T_k^{\mathscr{C}} \sigma XY = \sigma(T_k^{\mathscr{C}} X)(T_k^{\mathscr{C}} Y)$ y $G_k^{\mathscr{C}} \sigma XY = \sigma(G_k^{\mathscr{C}} X)(G_k^{\mathscr{C}} Y)$

   e) $T_k^{\mathscr{C}} \lambda X = \lambda T_k^{\mathscr{C}} X$ y $G_k^{\mathscr{C}} \lambda X = \lambda G_k^{\mathscr{C}} X$

3. para cada k tal que $0 \leq k < n-1$ las transformaciones naturales $\eta_k^{\mathscr{C}} : id \Rightarrow T_k^{\mathscr{C}}$ y $\varepsilon_k^{\mathscr{C}} : G_k^{\mathscr{C}} \Rightarrow id$ son *transformaciones SM* en el sentido de que cumplen las relaciones

$$\eta_k^{\mathscr{C}} \top = 1_{\top}, \qquad \varepsilon_k^{\mathscr{C}} \top = 1_{\top_o}, \qquad \eta_k^{\mathscr{C}}(X \otimes Y) = \eta_k^{\mathscr{C}} X \otimes \eta_k^{\mathscr{C}} Y,$$
$$\varepsilon_k(X \otimes Y) = \varepsilon_k^{\mathscr{C}} X \otimes \varepsilon_k^{\mathscr{C}} Y.$$

4. $(\mathscr{C}, T_k^{\mathscr{C}}, G_k^{\mathscr{C}}, \eta_k^{\mathscr{C}}, \varepsilon_k^{\mathscr{C}})$ es un modelo de $(\mathbf{n}, T_k, G_k, \eta_k, \varepsilon_k)$.

Teorema 4. *$(\mathscr{C}, T_k^{\mathscr{C}}, G_k^{\mathscr{C}}, \eta_k^{\mathscr{C}}, \varepsilon_k^{\mathscr{C}})$ es un modelo de $(\mathbf{n}, T_k, G_k, \eta_k, \varepsilon_k)$ si y sólo si la asignación*

$$T_k \rightsquigarrow T_k^{\mathscr{C}}$$
$$G_k \rightsquigarrow G_k^{\mathscr{C}}$$
$$\eta_k \rightsquigarrow \eta_k^{\mathscr{C}}$$
$$\varepsilon_k \rightsquigarrow \varepsilon_k^{\mathscr{C}}$$

es un 2-funtor SM.

Ejemplo 1. Si consideramos *Set* como categoría *SM* (en realidad *SM cartesiana* con el producto tensorial dado por $(\times, 1)$) podemos considerar el 2-funtor *SM*

$$(\mathbf{n}, T_k^n, G_k^n, \eta_k^n, \varepsilon_k^n) \longrightarrow (Set, T_k, G_k, \eta_k, \varepsilon_k)$$

como un modelo entre n-Comprensiones *SM*.

Definición 5. Sea $\mathscr{V}$ una categoría monoidal y $\mathscr{B}$ una $\mathscr{V}-$categoría.[4] Denominaremos *producto cotensorial de X y C*, y lo denotaremos por $X \multimap C$, a un objeto en $\mathscr{B}$ definido por el $\mathscr{V}-$isomorfismo natural

$$\mathscr{B}(B, X \multimap C) \equiv [X, \mathscr{B}(B,C)]$$

Si existe el producto cotensorial para todos X y C entonces la $\mathscr{V}-$categoría $\mathscr{B}$ se dice que es *cotensorial*.

Teorema 5. $\mathbf{n} \multimap \mathscr{C} \in \mathscr{SM}$ *para todo* $\mathscr{C} \in \mathscr{SM}$.

Proposición 3. *Siempre que el tuplo* $(\mathscr{C}, T_k, G_k, \eta_k, \varepsilon_k)$ *es una n-Comprensión* SM *podemos construir un tuplo* $(\mathbf{n} \multimap \mathscr{C}, \overline{T_k}, \overline{G_k}, \overline{\eta_k}, \overline{\varepsilon_k})$ *que también lo es.*

Definición 6. Sea X un objeto en una categoría SM $\mathscr{C}$. Llamamos *niveles de X* a las cadenas

$$\begin{tikzcd}
{[X_0] =} & {[X_1] =} & {[X_2] =} & \cdots & {[X_{n-1}] =} \\
X \arrow[d] & X \arrow[d, "id_X"] & X \arrow[d, "id_X"] & & X \arrow[d, "id_X"] \\
\top \arrow[d] & X \arrow[d] & X \arrow[d, "id_X"] & & X \arrow[d, "id_X"] \\
\top \arrow[d] & \top \arrow[d] & X \arrow[d] & & X \arrow[d, "id_X"] \\
\top \arrow[d] & \top \arrow[d] & \top \arrow[d] & & X \arrow[d, "id_X"] \\
\vdots & \vdots & \vdots & & \vdots \\
\top & \top & \top & & X
\end{tikzcd}$$

en la categoría $\mathbf{n} \multimap \mathscr{C}$.

Observación. Todos los objetos X_k con $k = 1, ..., n-1$ pueden generarse a partir de X_0 por medio de las ecuaciones $X_k = G_{k-1}X_{k-1}$ o bien, excluyendo X_{n-1}, por medio de $X_k = T_{k+1}X_{k+1}$ si $k = 0, ..., n-2$.

Definición 7. Definimos las doctrinas $\mathscr{A}^n$ como aquellas cuyos objetos son n-Comprensiones *SM* $(\mathscr{C}, T_k, G_k, \eta_k, \varepsilon_k)$ con $0 \leq k \leq n-2$

- dotadas de un diagrama unario (llamado *diagrama inicial*)

$$\top \xrightarrow{0} N_0 \xrightarrow{s} N_0$$

  en $\mathscr{C}$ tal que

$$T_0(\top \xrightarrow{0} N_0 \xrightarrow{s} N_0) = \top \xrightarrow{1_\top} \top \xrightarrow{1_\top} \top$$

[4] Véase (Kelly, 1989) para estos conceptos.

y además, si definimos recursivamente para cada $i = 1, 2, ..., n-2$ los objetos N_i por las reglas

$$N_1 = G_0 N_0$$

$$N_{i+1} = G_i N_i$$

tenemos para cada $i = 0, 1, ..., n-2$ y $j = 0, 1, ..., n-1$

$$T_i N_j = \begin{cases} \top & \text{si } i = j = 0 \\ N_{i-1} & \text{si } i = j \neq 0 \\ N_j & \text{en otro caso} \end{cases} \qquad G_i N_j = \begin{cases} N_{i+1} & \text{si } i = j \\ N_j & \text{en otro caso} \end{cases}$$

con $1 \leq j \leq n-2$. Con esto podemos generar todos los diagramas iniciales en la forma

$$\top \xrightarrow{0_j} N_j \xrightarrow{s_j} N_j$$

- cerradas bajo

  - *recursión plana (RP)*:

    para todos $g : X \longrightarrow Y$ y $h : N_0 \otimes X \longrightarrow Y$ donde $T_0 X$ y $T_0 Y$ son isomorfos a $\top$ para cada X e Y existe un único morfismo $f : N_0 \otimes X \longrightarrow Y$ en $\mathscr{C}$, que denotaremos por $RP(g, h)$, tal que el siguiente diagrama conmuta[5]

$$\begin{tikzcd} \top \otimes X \arrow[r, "0 \otimes X"] \arrow[dr, "g \circ \pi_1"'] & N_0 \otimes X \arrow[d, "f"] & N_0 \otimes X \arrow[l, "s \otimes X"'] \arrow[dl, "h"] \\ & Y & \end{tikzcd}$$

  - *recursión ramificada segura* en cada nivel k (RRS_k):

    para cada $k = 0, 1, ..., n-2$ y para todos los morfismos

$$g : X \longrightarrow Y \text{ y } h : Y \longrightarrow Y$$

    donde $T_k ... T_0 Y$ es isomorfo a $\top$ existe un único

$$f : N_{k+1} \otimes X \longrightarrow Y$$

[5]Este es realmente un diagrama de coproducto.

en $\mathscr{C}$ tal que conmuta el siguiente diagrama

$$\begin{tikzcd}
\top \otimes X \arrow[r, "0\otimes X"] \arrow[d, "i"'] & N_{k+1} \otimes X \arrow[r, "s\otimes X"] \arrow[d, "f"] & N_{k+1} \otimes X \arrow[d, "f"] \\
X \arrow[r, "g"'] & Y \arrow[r, "h"'] & Y
\end{tikzcd}$$

- y donde para cualquier par de objetos X e Y en $\mathscr{C}$ en la forma

$$\bigotimes_{i=0}^{n-1} N_i^{\alpha_i}$$

  el producto tensorial $X \otimes Y$ es también cartesiano (llamamos a estos objetos *cartesianos*)

- para los objetos cartesianos las n-Comprensiones $(\mathscr{C}, T_k, G_k, \eta_k, \varepsilon_k)$ también están cerradas bajo el esquema de *recursión dependiente segura* en cada nivel k (RDS_k):

  para cada $k = 0, 1, ..., n-2$ y para todos los morfismos

$$g : X \longrightarrow Y \text{ y } h : (N_{k+1} \otimes X) \otimes Y \longrightarrow Y$$

  donde $T_k ... T_0 Y$ es isomorfo a $\top$ y X e Y son objetos cartesianos existe un único

$$f : N_{k+1} \otimes X \longrightarrow Y$$

  en $\mathscr{C}$ tal que conmuta el siguiente diagrama

$$\begin{tikzcd}
\top \otimes X \arrow[r, "0_{k+1}\otimes X"] \arrow[dr, "(0_{k+1}\otimes X), g\circ\pi_1"'] & N_{k+1} \otimes X \arrow[r, "s_{k+1}\otimes X"] \arrow[d, "id, f"] & N_{k+1} \otimes X \arrow[d, "f"] \\
& (N_{k+1} \otimes X) \otimes Y \arrow[r, "h"'] & Y
\end{tikzcd}$$

Los objetos en la doctrina $\mathscr{A}^n$ son entonces *n-Comprensiones SM* con cuatro tipos de diagramas diferentes y unas condiciones de acotación respecto a los objetos sobre los que éstos diagramas actúan.

Definición 8. Sean $\mathbb{N}_k = \{(k,n)/n \in \mathbb{N}\}$ con $k = 0, 1, ..., n-2$ los conjuntos tales que cada par (k,n) puede asociarse con un número natural n_k en un nivel diferente según la Definición 6.

De este modo podemos escribir $s_k(k,n) = (k, n+1)$ para cada $n \in \mathbb{N}$ y cada $k = 0, 1, ..., n-2$

donde s_k es la función sucesor en el nivel k del *diagrama inicial* en *Set* dado por

$$1 \xrightarrow{0} \mathbb{N}_0 \xrightarrow{s} \mathbb{N}_0$$

de acuerdo con la Definición 7 y todas las ecuaciones que hacen de $(Set, T_k, G_k, \eta_k, \varepsilon_k)$ una n-Comprensión *SM*.

Proposición 4. *La doctrina $\mathscr{A}^n$ está dotada de una categoría inicial.*

Denotaremos por $\mathscr{J}^n$ a la categoría inicial en la doctrina $\mathscr{A}^n$.

Definición 9. Una *categoría simétrica monoidal cartesiana* (*categoría SM cartesiana* en adelante) es una categoría simétrica monoidal cuya estructura monoidal viene dada por el producto cartesiano.

Observación. De la Definición anterior se extrae el hecho de que la unidad del producto tensorial es un objeto terminal de la categoría.

Proposición 5. *Toda categoría SM cartesiana está dotada para cada objeto X de* morfismos duplicación *en la forma $\delta_X : X \to X \otimes X$ y* morfismos de borrado *en la forma $\tau_X : X \to \top$.*

En la categoría $\mathscr{J}^n$ podemos definir los morfismos duplicación y borrado de manera que se satisfagan las hipótesis de la Proposición anterior.

Los morfismos de borrado $\tau_X : X \longrightarrow \top$ en $\mathscr{J}^n$ se definen recursivamente considerando:

1. si $X = \top$ tomamos $\tau_\top = 1_\top$
2. si $X = N_0$ utilizamos la composición $\tau_{N_0} = d \circ T_0 \circ d^{-1} : N_0 \longrightarrow N_0 \otimes \top \longrightarrow \top \otimes \top \longrightarrow \top$
3. si $X = N_k$ con $k = 1, ..., n-2$

$$\begin{tikzcd}
\top \otimes \top \arrow[r, "0_{k+1} \otimes \top"] \arrow[dr, "i"'] & N_{k+1} \otimes \top \arrow[r, "s_{k+1} \otimes \top"'] \arrow[d, "f"] & N_{k+1} \otimes \top \arrow[d, "f"] \\
& \top \arrow[r] & \top
\end{tikzcd}$$

   y la composición $\tau_{N_k} = f \circ d^{-1}$

4. si $X = Y \otimes Z$ con Y y Z en alguno de los casos anteriores entonces tenemos igualmente el morfismo de borrado recordando que $\tau_X X = \tau_X Y \otimes \tau_X Z$

De esta manera obtenemos los morfismos de borrado en cualquier objeto de $\mathscr{J}^n$.

Los morfismos de duplicación δ_{N_k} se obtienen por los diagramas

$$\begin{tikzcd}
\top \otimes \top \arrow[r, "0_{k+1}\otimes\top"] \arrow[dr, "0_k\otimes 0_k"'] & N_{k+1}\otimes\top \arrow[r, "s_{k+1}\otimes\top"] \arrow[d, "f"] & N_{k+1}\otimes\top \arrow[d, "f"] \\
& N_k\otimes N_k \arrow[r, "s_k\otimes s_k"'] & N_k\otimes N_k
\end{tikzcd}$$

para todo $k = 1, ..., n-2$ y la composición $\delta_{N_k} = G_k(f \circ d^{-1})$.

El caso de δ_{N_0} tiene el problema de que no disponemos ni de un diagrama que nos lo proporcione ni unos funtores de coerción que nos permitan, cuando $n = 2$, bajar un nivel el objeto sobre el que actúa. Consideraremos por tanto en lo sucesivo la categoría inicial $\mathscr{J}^n$ siempre con $n > 2$.

Definición 10. El *modelo estandar de los morfismos formales* es el funtor Γ_n expresado por el diagrama

$$\begin{tikzcd}
\mathscr{J}^n \arrow[rr, "\Gamma_n"] \arrow[dr, "\overline{\chi}"'] & & n \multimap Set \\
& n \multimap \mathscr{J}^n \arrow[ur, "n\multimap\Gamma"'] &
\end{tikzcd}$$

esto es $\Gamma_n = (n \multimap \Gamma) \circ \overline{\chi}$ donde Γ es el funtor $\Gamma : \mathscr{J}^n \longrightarrow Set$ definido por $\Gamma X = \mathscr{J}^n(\top, X)$ y $\Gamma f = f \circ -$.

Se va a presentar a continuación la caracterización de las clases $\mathscr{E}^n$ por medio de las n-doctrinas iniciales $\mathscr{J}^n$ con $n > 2$.

Como se ha señalado en la Introducción, en $\mathscr{J}^3$ se produce la situación de que la especie de las variables de los morfismos puede ser llevada hasta el cero por medio de endofuntores T_0 y T_1. Ello provoca que todas las funciones numéricas definibles en $\mathscr{J}^3$ cuyo dominio y codominio no contenga N_2 pueden reducirse a funciones cuyo dominio y codominio sean productos tensoriales de N_0, lo cual obviamente nos situaría fuera de cualquier jerarquía de funciones subrecursivas de crecimiento progresivo tal como se han presentado al principio.

Pasamos a continuación por tanto a definir una clase $\mathscr{K}^n$ de flechas de $\mathscr{J}^n$ sobre las que no permitiremos la actuación de los endofuntores T_k. Su descripción por lo tanto es la misma que la de $\mathscr{J}^n$ excepto que en el nivel $k+1$ introducimos dos cambios:[6]

1. cambiamos las funciones generadas por *RRS* por las funciones generadas por *RDS* y

[6]Obsérvese que $\mathscr{K}^n$ ya no satisfacen las condiciones que harían de ellas unas doctrinas en $\mathscr{A}^n$.

2. en dicho nivel solamente consideramos las flechas que surgen de las aplicaciones de los endofuntores *id* y G_k sobre los morfismos del nivel k .

Los objetos de $\mathscr{J}^n$ y $\mathscr{K}^n$ son los mismos.

Se va a definir, siguiendo (Wirz, Wirz), las clases $\mathscr{B}^n$. Para ello usaremos un lenguaje que contiene n *especies de variables* diferentes, separadas por puntos y coma, que denotaremos por los números $0, 1, ..., n-1$.

Definición 11. Para cada $n \geq 0$ definimos la clase $\mathscr{B}^{n+1}$ como la más pequeña de las clases de funciones conteniendo las funciones iniciales

1. (constante cero) 0 de nivel 0
2. *(proyecciones)* $\pi_{k,j}(x_{n,1}, ..., x_{n,r_n}; ...; x_{0,1}, ..., x_{0,r_0}) = x_{k,j}$ para cada $0 \leq k \leq n$ y $1 \leq j \leq r_k$ de nivel k
3. (*sucesor*) $S(x_0) = x_0 + 1$ de nivel 0
4. (*predecesor*) $P(0_0) = 0$, $P(x_0 + 1) = x_0$ de nivel 0
5. (*condicional*) $C(0_0, y_0, z_0) = y_0$, $C(x_0 + 1, y_0, z_0) = z_0$ de nivel 0

y está cerrada bajo las operaciones

1. (*recursión segura*)

$$\begin{cases} f(\overline{x_n}; ...; \overline{x_{k+2}}; \overline{x_{k+1}}, 0_{k+1}; \overline{x_k}) = g(\overline{x_n}; ...; \overline{x_{k+2}}; \overline{x_{k+1}}; \overline{x_k}) \\ f(\overline{x_n}; ...; \overline{x_{k+2}}; \overline{x_{k+1}}, v_{k+1} + 1; \overline{x_k}) = \\ h(\overline{x_n}; ...; \overline{x_{k+2}}; \overline{x_{k+1}}, v_{k+1}; \overline{x_k}, f(\overline{x_n}; ...; \overline{x_{k+2}}; \overline{x_{k+1}}, v_{k+1}; \overline{x_k})) \end{cases}$$

   donde g y h están en $\mathscr{B}^{n+1}$ y el nivel de f es el máximo de los niveles de g y h

2. (*composición segura*)

$$f(\overline{x_n}; ...; \overline{x_0}) = h(\overline{r_n}(\overline{x_n};); \overline{r_{n-1}}(\overline{x_n}; \overline{x_{n-1}}); ...; \overline{r_0}(\overline{x_n}; ...; \overline{x_0}))$$

   donde $\overline{r_0}, ..., \overline{r_n}$ y h están en $\mathscr{B}^{n+1}$ y el nivel de f es el nivel de h mientras que el nivel de r_n es menor o igual que n y el de r_0 es 0.

De acuerdo con (Wirz, Wirz), que generaliza (Bellantoni & Cook, 1992), las clases $\mathscr{B}^n$ dan lugar a otras que resultan ser equivalentes a las clases de funciones $\mathscr{E}^n$ que forman la *Jerarquía de Grzegorzcyk*. Las nuevas clases se obtienen considerando aquellas funciones de $\mathscr{B}^n$ cuyas variables se han colocado de forma que son de la misma especie.

Definición 12. Sean $(\mathscr{B}^n)^\star$ las clases de funciones que coinciden con $\mathscr{B}^n$ considerando que todas sus variables son de la misma especie.

Teorema 6. $(\mathscr{B}^n)^\star = \mathscr{E}^n$ *para cada* $n \in \mathbb{N}$.

Se va a presentar en la presente Sección la caracterización de las clases $\mathscr{E}^n$ por medio de las clases $\mathscr{K}^n$ con $n > 2$.

Teorema 7. *Sean*

$$f^\flat : \prod_{i=0}^{n-1} \mathbb{N}_i^{\alpha_i} \longrightarrow \mathbb{N}_j \ , \ f^\sharp : \mathbb{N}_{n-1}^{\sum_{i=0}^{n-1}\alpha_i} \longrightarrow \mathbb{N}_{n-1} \ , \ f : \mathbb{N}^{\sum_{i=0}^{n-1}\alpha_i} \longrightarrow \mathbb{N}$$

las funciones tales que, si identificamos los números n_k *(*$k = 0, \ldots, n-1$*) con* n*, las funciones* f, f^b *y* $f^\sharp$ *dan el mismo valor para los argumentos apropiados que sean iguales en la identificación anterior. Entonces*

$$f^b \in \mathscr{B}^n \longrightarrow f^\sharp \in \mathscr{B}^n \longleftrightarrow f \in (\mathscr{B}^n)^\star$$

Definición 13. Denominamos $\mathscr{G}^n$ al conjunto de funciones[7]

$$\{\Gamma_n T_{n-2} \ldots T_0 f / f \in \mathscr{K}^n \text{ y } cod(f) = N_{n-1}\}$$

Teorema 8. $\mathscr{G}^n = \{\Gamma_n f : \mathbb{N}_{n-1}^{\alpha} \longrightarrow \mathbb{N}_{n-1} / f \in \mathscr{K}^n\}$.

Antes de pasar a la demostración del Teorema principal recordamos que cada función f de $\mathscr{B}^n$ tiene la forma

$$f(\overline{x_{n-1}}; \overline{x_{n-2}}; \ldots; \overline{x_0})$$

de lo que se deduce, ordenando al revés sus argumentos, que tiene la forma

$$f : \prod_{i=0}^{n-1} \mathbb{N}_i^{\alpha_i} \longrightarrow \mathbb{N}_{n-1}$$

donde algunos α_j pueden ser ceros.[8]

[7] $cod(f)$ significa aquí codominio de la función f.
[8] Suponemos que tanto el producto cartesiano como el tensorial asocian por la derecha.

Teorema 9. *$\mathscr{B}^n = \mathscr{G}^n$ para todo $n > 2$.*

Demostración. Probaremos las dos inclusiones basándonos en (Wirz, Wirz).

1. $\boxed{\mathscr{B}^n \subseteq \mathscr{G}^n}$

   Vamos a demostrar en su lugar que para cada función $f \in \mathscr{B}^n$ tal que

$$f : \prod_{i=0}^{n-1} \mathbb{N}_i^{\alpha_i} \longrightarrow \mathbb{N}_k$$

   hay una f' en $\mathscr{K}^n$ tal que

$$f' : \bigotimes_{i=0}^{n-1} N_i^{\alpha_i} \longrightarrow N_k$$

   Además f y f' toman los mismos valores aplicados a los argumentos adecuados (se considera que los números en $\mathbb{N}_j$ se corresponden con los de N_j). Diremos que f y f' *concuerdan*. Esto lo demostramos por inducción sobre la construcción o definición de la función f en $\mathscr{B}^n$.

   - Si f es una función inicial tenemos los casos

     a) f es la función cero: $0 : 1 \longrightarrow \mathbb{N}_j$ concuerda con

$$0_j : \top \longrightarrow N_j$$

     b) f es la función sucesor: $s : \mathbb{N}_j \longrightarrow \mathbb{N}_j$ concuerda con

$$s_j : N_j \longrightarrow N_j$$

     c) la proyección $\pi_{k,j} : \prod_{i=0}^{n-1} \mathbb{N}_i^{\alpha_i} \longrightarrow \mathbb{N}_k$ concuerda con la función[9]

$$\bigotimes_{i=0}^{k-1} \tau_i^{\alpha_i} \otimes \bigotimes_{i=1}^{j-1} \tau_k \otimes 1_{N_k} \otimes \bigotimes_{i=j+1}^{\alpha_k - j} \tau_k \otimes \bigotimes_{i=k+1}^{n-1} \tau_i^{\alpha_i}$$

     d) f es la función predecesor. Concuerda con la función

$$RP(0, 1_{N_0}) : N_0 \longrightarrow N_0$$

[9] Esta función la denotaremos por $\pi'_{k,j}$.

e) f es la función condicional. Concuerda con la función

$$RP(\pi_0 \circ i, \pi_1 \circ \pi_1) : N_0 \otimes (N_0 \otimes N_0) \longrightarrow N_0$$

- f se obtiene por composición segura en la forma

$$f(\overline{x_{n-1}}; \ldots ; \overline{x_0}) = h(\overline{r_{n-1}}(\overline{x_n}; \overline{x_{n-1}};); \ldots ; \overline{r_0}(\overline{x_n}; \overline{x_{n-1}}; \ldots ; \overline{x_0}))$$

Así tendremos $h : \prod_{i=0}^{n-1} \mathbb{N}_i^{\alpha_i} \longrightarrow \mathbb{N}_j$ para cada $s = 0, \ldots, n-1$ una sucesión de funciones

$$\{r_{s,t} : \prod_{l=s}^{n-1} \mathbb{N}_l^{\beta_l} \longrightarrow \mathbb{N}_s\}_{1 \leq t \leq \alpha_s}$$

y $f : \prod_{1=0}^{n-1} \mathbb{N}_i^{\beta_i} \longrightarrow \mathbb{N}_m$. Para simplificar la demostración vamos a considerar que

$$h : \mathbb{N}_{n-1} \times \mathbb{N}_{n-2}^2 \longrightarrow \mathbb{N}_j$$

$$r : \mathbb{N}_{n-1}^3 \longrightarrow \mathbb{N}_{n-1}$$

$$t : \mathbb{N}_{n-1}^3 \times \mathbb{N}_{n-2}^2 \longrightarrow \mathbb{N}_{n-2}$$

$$s : \mathbb{N}_{n-1}^3 \times \mathbb{N}_{n-2}^2 \longrightarrow \mathbb{N}_{n-2}$$

Entonces $f : \mathbb{N}_{n-1}^3 \times \mathbb{N}_{n-2}^2 \longrightarrow \mathbb{N}_j$. Por hipótesis de inducción suponemos que h, r, t, s concuerdan con las funciones h', r', t', s' entonces f concuerda con la siguiente composición

$$\mathbb{N}_{n-1}^3 \times \mathbb{N}_{n-2}^2 \xrightarrow{r' \circ p_{0,3}, t', s'} \mathbb{N}_{n-1} \times \mathbb{N}_{n-2}^2 \xrightarrow{h'} \mathbb{N}_j$$

donde $p_{e,d}$ con $e < d$ es la función cuyo dominio tiene la forma $\bigotimes_{i=0}^{k} X_i$ para $d \leq k$, su codominio es $\bigotimes_{i=e}^{d} X_i$ y es igual a

$$a^{d-e} \pi_0 \circ a^{-d+e} \circ \overbrace{\pi_1 \circ \pi_1}^{e \text{ veces}}$$

- f se obtiene por recursión segura a partir de $g, h \in \mathscr{B}^n$:

$$f(\overline{x_{n-1}}; \ldots ; \overline{x_{k+2}}; \overline{x_{k+1}}, 0_{k+1}; \overline{x_k}) = g(\overline{x_{n-1}}; \ldots ; \overline{x_{k+2}}; \overline{x_{k+1}}; \overline{x_k})$$

$$f(\overline{x_{n-1}}; \ldots ; \overline{x_{k+2}}; \overline{x_{k+1}}, z_{k+1} + 1; \overline{x_k}) =$$

$$h(\overline{x_{n-1}};\ldots;\overline{x_{k+2}};\overline{x_{k+1}},v_{k+1};\overline{x_k},f(\overline{x_{n-1}};\ldots;\overline{x_{k+2}};\overline{x_{k+1}},z_{k+1};\overline{x_k}))$$

por hipótesis de inducción g y h concuerdan cambiando el orden de los argumentos con las funciones g' y h' en $\mathscr{K}^n$

$$g' : \bigotimes_{i=k+1}^{n-1} N_i^{\alpha_i} \otimes N_k^{\alpha_k} \longrightarrow N_k$$

$$h' : N_{k+1} \otimes \bigotimes_{i=k+1}^{n-1} N_i^{\alpha_i} \otimes N_k^{\alpha_k} \otimes N_k \longrightarrow N_k$$

Sea $X = \bigotimes_{i=k}^{n-1} N_i^{\alpha_i}$ e $Y = N_k$, es fácil ver que la única función f' que existe por el esquema *RDS*

$$\begin{tikzcd}
\top \otimes X \arrow[r, "0\otimes X"] \arrow[dr, "(0\otimes X),g'\circ\pi_1"'] & N_{k+1}\otimes X \arrow[r, "s\otimes X"] \arrow[d, "id,f'"] & N_{k+1}\otimes X \arrow[d, "f'"] \\
& (N_{k+1}\otimes X)\otimes Y \arrow[r, "h'"'] & Y
\end{tikzcd}$$

concuerda cambiando el orden de los argumentos con la función f.

2. $\boxed{\mathscr{G}^n \subseteq \mathscr{B}^n}$

Vamos a demostrar que con cada función $f \in \mathscr{K}^n$ tal que

$$cod(f) = \bigotimes_{i=0}^{n-1} N_i^{\beta_i}$$

para cada $j \in n$ con $\beta_j \neq 0$ las β_j funciones $f_{j,k}$ con $k = 1,\ldots,\beta_j$, a las que llamaremos *proyecciones de* f, tales que

$$f_{j,k} = \pi'_{j,k} \circ f$$

hay una función en $\mathscr{B}^n$ que concuerda con $f_{j,k}$. La demostración se hace por inducción en el nivel de las funciones en $\mathscr{K}^n$.

- f es de nivel 0:

  a) f es la función cero o sucesor. Concuerdan con las correspondientes en $\mathscr{B}^n$.

  b) f es la función $\varepsilon_k(N_k) : N_{k+1} \longrightarrow N_k$ que concuerda con la función definida por la recursión segura

$$\begin{cases} f(0;\ldots;) = 0 \\ f(n+1;\ldots;) = s(;f(n;\ldots;)) \end{cases}$$

al elevar la especie del valor de f.

- f se obtiene por producto tensorial de g y h. Las proyecciones de f son g o h o las proyecciones de g o las de h.
- f surge de los isomorfismos naturales a, s o d, las proyecciones de f concuerdan con alguna de las proyecciones $\pi_{j,k}$.
- f es la composición $h \circ g$. Para simplificar consideremos que

$$g : N_j \longrightarrow N_s \otimes N_t \qquad y \qquad h : N_s \otimes N_t \longrightarrow N_i \otimes N_k \otimes N_u$$

Entonces $f : N_j \longrightarrow N_i \otimes N_k \otimes N_u$. Por hipótesis, con las dos proyecciones de g, g_1 y g_2, y con las tres de h, h_1, h_2 y h_3, existen las funciones g'_1, g'_2, h'_1, h'_2, h'_3 que concuerdan con ellas. Vamos a ver que con cada proyección de f, f_1, f_2 y f_3, hay una función en $\mathscr{B}^n$ que concuerda con ella.

$$f'_1(x_j;) = h'_1(g'_1(x_j); g'_2(x_j))$$

$$f'_2(x_j;) = h'_2(g'_1(x_j); g'_2(x_j))$$

$$f'_3(x_j;) = h'_3(g'_1(x_j); g'_2(x_j))$$

- f se obtiene por recursión plana a partir de

$$g : X \longrightarrow Y \text{ y } h : N_0 \otimes X \longrightarrow Y$$

donde $Y = N_0^k$ (eliminamos el caso en que $Y = \top$ que no da lugar a funciones numéricas y el caso en que uno de sus factores en el producto tensorial sea $\top$):

$$\begin{tikzcd}
\top \otimes X \arrow[r, "0\otimes X"] \arrow[dr, "g\circ i"'] & N_0 \otimes X \arrow[d, "f"] & N_0 \otimes X \arrow[l, "s\otimes X"'] \arrow[dl, "h"] \\
& Y &
\end{tikzcd}$$

entonces tenemos

$$f \circ (0 \otimes X) = g \circ i$$

$$f \circ (s \otimes X) = h$$

Por hipótesis de inducción con cada proyección de g, g_i $(j = 1, \ldots, k)$ concuerda una función $g'_j \in \mathscr{B}^n$ y con cada proyección de h, h_j $(j = 1, \ldots, k)$ concuerda una función

$h'_j \in \mathscr{B}^n$. Es fácil ver que si f_j $(j = 1, \ldots, k)$ son las proyecciones de f, tenemos

$$f_j \circ (0 \otimes X) = g_j \circ i$$

$$f_j \circ (s \otimes X) = h_j$$

Así tenemos que con f_j concuerda la función f'_j definida por la recursión segura que sigue

$$f'_j(0, \bar{x}) = g'_j(\bar{x})$$

$$f'_j(n+1, \bar{x}) = h'_j(\bar{x})$$

- f se obtiene por recursión dependiente segura de nivel k a partir de las funciones $g : X \longrightarrow Y$ y $h : N_{k+1} \otimes X \otimes Y \longrightarrow Y$ (suponemos que Y es el producto tensorial de k factores de la forma N_s con $s < k+1$):

$$\begin{tikzcd}
\top \otimes X \arrow[r, "0\otimes X"] \arrow[dr, "(0\otimes X), g\circ\pi_1"'] & N_{k+1} \otimes X \arrow[r, "s\otimes X"] \arrow[d, "id, f"] & N_{k+1} \otimes X \arrow[d, "f"] \\
& (N_{k+1} \otimes X) \otimes Y \arrow[r, "h"'] & Y
\end{tikzcd}$$

Por hipótesis de inducción con cada proyección g_i de g concuerda una función g'_i y con cada proyección h_i de h concuerda una función h'_i. Es fácil ver que con cada proyección f_i de f concuerda la función f'_i. Las funciones f'_i se definen por recursión simultánea donde no ponemos las especies de las variables, pero que es fácil ver que se cumplen los requisitos de la recursión segura:

$$\begin{aligned}
& f'_i(0, \bar{x}) = g'_i(\bar{x}) \text{ para } i = 1, \ldots, s \\
& f'_1(n+1, \bar{x}) = h'_i(n, \bar{x}, f'_1(n, \bar{x}), \ldots, f'_s(n, \bar{x})) \\
& f'_2(n+1, \bar{x}) = h'_2(n, \bar{x}, f'_1(n, \bar{x}), \ldots, f'_s(n, \bar{x})) \\
& \qquad \ldots\ldots\ldots \\
& f'_s(n+1, \bar{x}) = h'_s(n, \bar{x}, f'_1(n, \bar{x}), \ldots, f'_s(n, \bar{x}))
\end{aligned}$$

Análogamente a como se demuestra que las funciones recursivas primitivas están cerradas respecto a la recursión simultánea, puede demostrarse que $\mathscr{B}^3$ está cerrada respecto a la recursión segura simultánea, ya que puede reducirse a la segura usando la función

$$\varphi(n, \bar{x}) = \prod_{i=1}^{k} p_i^{f'_i(n, \bar{x})}$$

que pertenece a $\mathscr{B}^n$ si las f_i' pertenecen a $\mathscr{B}^n$.

□

REFERENCES

Bellantoni, S. & Cook, S. (1992). New recursion-theoretic characterization of the polytime functions. *Comput. Complexity*, *2*, 91–110.

Burroni, A. (1986). Récursivité graphique (1e partie):catégorie des fonctions récursives primitives formelles. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, *27*(1), 49–79.

Cockett, R. & Redmond, B. (2010). A categorical setting for lower complexity. *Electronic Notes in Theoretical Computer Science*, *265*, 277–300.

Kelly, G. (1989). Elementary observations on 2-categorical limits. *Bulletin of the Australian Mathematical Society*, *39*, 301–317.

Lambek, J. & Scott., P. (1986). *Introduction to Higher Order Categorical Logic*. CUP.

Leivant, D. (1994). *Feasible Mathematics II*, chapter Ramified recurrence and computational complexity I: Word algebras and polytime. Birkhause.

Otto, J. (1995). *Complexity Doctrines.* PhD thesis, McGill University.

Roman, L. (1989). Cartesian categories with natural numbers object. *Journal of Pure and Applied Algebra*, *28*, 267–278.

Roman, L. (1990). Categories with finite limits and natural numbers object. Technical report, Publicaciones preliminares del Instituto de Matemáticas.

Thibault, M.-F. (1977). *Représentations des fonctions récursives dans les catégories*. PhD thesis, McGill University.

Wirz, M. Characterizing the grzegorczyk hierarchy by safe recursion. Retrieved on 2009-04-21.

ABDUCCIÓN Y DERECHO

EMILIO GARCÍA BUENDÍA
Universidad Complutense de Madrid
Departamento de Lógica y Filosofía de la Ciencia
egarciabue@yahoo.es

ABSTRACT: This paper pursues a double purpose. The first objective shows the general characteristics of the abductive reasoning in order to explain both, similarities and differences, with the deductive and inductive ones. The second objective shows the potential applications of this kind of reasoning in the field of Legal Practice in order to suggest that, to achieve the desired results on this matter, the way forward should be the integration of basically four types of inferences: deductive, inductive, abductive and analogical.
KEYWORDS: Abduction, Abductive reasoning, Law reasoning, Legal logic, Artificial Intelligence and Law

1. El razonamiento abductivo en general

El razonamiento abductivo está volviendo a despertar gran atención tanto en las ciencias cognitivas como en la inteligencia artificial si bien aquí se considera que puede tener una gran relevancia en el campo del razonamiento jurídico.

Hoy en día se encuentra plenamente aceptado que, en las distintas actividades humanas y vinculado directamente con la racionalidad, existen varias formas de razonamientos que captan las argumentaciones tanto del sentido común como a la hora de explicar contradicciones aparentes que se nos presentan como en la esfera del descubrimiento científico y su progreso. Dichas formas de razonamientos serían: la deductiva, la inductiva, la abductiva y la analógica. Dentro de cada una de las anteriores se pueden realizar diversas subdivisiones cada una de ellas con sus propias características.

Centrándonos en el razonamiento abductivo se han puesto una gran diversidad de ejemplos siendo el de la práctica médica a la hora de diagnosticar uno de los más típicos. Cuando un médico observa unos síntomas en un paciente busca, entre diversas hipótesis plausibles, aquella de todas ellas que pueda explicar o dar cuenta de la totalidad de los síntomas observados teniendo en cuenta una base de conocimiento de carácter previo que le suministra una información acerca de las relaciones causales entre las enfermedades y síntomas.

La inferencia abductiva, a groso modo, es aquella que concluye en una hipótesis explicativa de un hecho extraño (en palabras de Peirce) a partir de dicho hecho y un cierto conocimiento previo

o base de conocimiento[1]. Por ello, la abducción es una *lógica* de la generación de hipótesis explicativas, o del "descubrimiento", un proceso de razonamiento invocado para explicar una observación sorprendente.

Los trabajos hasta la fecha sobre este tipo de razonamiento han prestado mayor atención a la abducción verbal, es decir, al razonamiento abductivo con representaciones de información de tipo lingüístico. Existen muchas conclusiones que no se alcanzan de manera deductiva tanto en diversos campos de conocimientos como en la vida diaria. De hecho, se considera que muchas conclusiones se alcanzan probabilísticamente en términos de aplicación de esquemas que ajustan un fenómeno dentro de un patrón determinado sin emplear un razonamiento deductivo denominándose comúnmente como "inferencia de la mejor explicación".

Como otro ejemplo de este tipo de razonamiento se puede considerar el siguiente: pensemos que vamos por la playa y vemos una imagen de Churchill dibujada sobre la arena. Cabe la posibilidad que dicho dibujo lo hubiera hecho una hormiga que caminó sobre la arena tal y como plantea Putnan en su obra *Razón, verdad e historia* pero la conclusión que sacamos es que alguien ha pintado dicha imagen sobre la arena. Dicha conclusión no se deriva necesariamente de las premisas si bien se considera como la explicación más probable a un hecho concreto.

Si BC es una base de conocimiento previo, y h es un hecho sin explicación, hay abducción si se puede explicar dicho hecho h a partir de la conjunción de dicha base de conocimiento (BC) y una hipótesis (H) siendo consistentes BC y H.

Es por ese motivo que al razonamiento abductivo se le ha denominado *deducción a la inversa* dado que la conclusión ya se encuentra presente y lo que se necesita encontrar es una premisa hipotética que, siendo consistente con el resto de las premisas, permita derivar la conclusión o hecho "sorprendente" o hecho a explicar.

Tomando los símbolos anteriores, de BC y h se abduce H, si y solo si h es consecuencia lógica de BC y H. En eso consiste la reversibilidad entre deducción y abducción. Desde un punto de vista formal se representaría del siguiente modo:

$$\{BC \wedge ?\} \models h \qquad\qquad \{BC \wedge H\} \models h$$

[1] Muchos autores denominan a dicha base de conocimientos "teoría de base" si bien no se comparte dicha terminología por considerar que una teoría tiene otra caracterización muy distinta.

Por lo anterior, siempre se ha creído que lo contrario de la deducción es la inducción si bien este esquema argumentativo parece demostrar justamente lo contrario, es decir, que lo opuesto a la deducción sería la abducción.

El hecho de considerar al razonamiento abductivo como la otra cara de la deducción disponiendo de una lógica susceptible de ser sometida a las diversas herramientas de dicha disciplina hace que sea representable en programación lógica, o lo que es lo mismo, en IA con el fin de emular razonamientos reales.

A fin de diferenciar la inferencia abductiva de la inductiva se puede considerar que ambas son ampliativas en el sentido de considerar que la conclusión va más allá de lo que se deduce lógicamente de las premisas si bien en la inductiva únicamente se tienen en cuenta frecuencias de observaciones determinadas; en la inferencia abductiva se atiende principalmente a consideraciones explicativas.

Por otro lado, normalmente en la inferencia inductiva se busca encontrar una enunciado universal mientras que en la inferencia abductiva la conclusión se encuentra formada por un enunciado particular que recoge un hecho singular necesitado de explicación.

Otra característica del razonamiento abductivo que compararte con el inductivo pero no así con el deductivo dentro de la lógica clásica es que no tiene la propiedad de la monotonía. Esto significa que cabe la posibilidad de deducir abductivamente una conclusión a partir de un subconjunto de premisas del conjunto β sin que se pueda llevar a cabo a partir del conjunto total. En el ejemplo anterior, si se añade otra premisa, como por ejemplo la que afirme que en esa playa finalizó hace tres horas un concurso de dibujos sobre la arena, puede conducir a otro tipo de conclusión que podría ser el considerar que dicha figura fue la ganadora de aquel concurso. Lo importante radica en determinar cuál es la mejor explicación posible la cual puede variar en función de la ampliación de los supuestos.

El razonamiento abductivo se ha dividido en abducciones selectivas y abducciones creativas; las primeras son aquellas que seleccionan la mejor conclusión dentro de un conjunto de posibles explicaciones mientras que las segundas introducirían nuevos patrones explicativos o conceptuales.

Pero es importante igualmente llamar la atención sobre la distinción entre los llamados “razonamientos ideales” frente a los “razonamientos reales”. Si una inferencia deductiva garantiza que la conclusión se sigue necesariamente de las premisas cabe plantearse si esto se puede encontrar en el mundo real o queda reducido únicamente al mundo ideal de las ciencias formales. Si nos situamos en el mundo real y en las inferencias que se llevan a cabo en vida cotidiana rápidamente nos damos cuenta de dicha necesariedad propia de la deducción nunca se da. Por el contrario, el

viada práctica aparecen razonamientos de tipo probabilístico, aproximativo o aquellos en los que se buscan explicaciones a hechos observados a nuestro alrededor. Es ahí donde el razonamiento abductivo aparece como una de las principales formas de dichos "razonamientos reales".

Y uno de los campos más humanos y reales por excelencia es precisamente el relacionado con el Derecho y su aplicación práctica.

En el campo jurídico, los modelos que se han desarrollado, en gran medida, se han basado en diversas extensiones de las lógicas modales precisamente dentro de ese mundo "ideal" ya mencionado. Los diferentes sistemas de lógicas deónticas se sitúan dentro del ámbito del razonamiento deductivo lo cual podría ser una de las causas de sus escasos éxitos por la razón antes expuesta entre otras, es decir, debido a su escaso contacto con la realidad. Dicho de otro modo, cabe la posibilidad de que el razonamiento deductivo no permita captar los procesos inferenciales que se llevan a cabo en el campo de la aplicación del Derecho o al menos en muchas de sus áreas..

La actividad jurisdiccional se ejerce en una gran pluralidad de supuestos dentro de lo que se puede denominar "teorías de la evidencia" o "teorías del caso" en las que se lleva a cabo una abducción que genera la resolución basándose en que dicha conclusión es la mejor explicación ante la evidencia existente. Habitualmente se parte de hechos o afirmaciones sobre estados de cosas que presentan cierto grado de inconsistencias siendo preciso elaborar una abducción para encontrar la mencionada "mejor explicación" posible que dé cuenta no solo de las pruebas presentadas sino de las inconsistencias que aparecen entre ellas *prima facie*.

2. El razonamiento abductivo en el Derecho penal.

Si se intenta mostrar una posible aplicación de esquema inferencial abductivo dentro del campo del Derecho Penal, lo primero que será preciso será acudir a la ya denominada "Base de Conocimiento" la cual vendría suministrada por los textos legales dado su carácter normativo y fuente del Derecho.

Siguiendo con este caso, y limitándonos a la cuestión de la autoría de la responsabilidad penal, el Código Penal, en su artículo 28 prescribe lo siguiente:

> "Son autores quienes realizan el hecho por sí solos, conjuntamente o por medio de otro del que se sirven como instrumento.
>
> También serán considerados autores:
>
> Los que inducen directamente a otro u otros a ejecutarlo.
> Los que cooperan a su ejecución con un acto sin el cual no se habría efectuado."

El anterior precepto penal recogería el conjunto de enunciados que constituirían la Base de Conocimiento (BC). Dichos enunciados serían los siguientes:

Enunciado BC1 (E_{BC1}) = "Persona que realiza el hecho por sí sólo" (art. 28 par. 1CP)
Enunciado BC2 (E_{BC2}) = "Persona que realiza el hecho conjuntamente" (art. 28 par. 1CP)
Enunciado BC3 (E_{BC3}) = "Persona que realiza el hecho por medio de otro" (art. 28 par. 1CP)
Enunciado BC4 (E_{BC4}) = "Persona que induce a otro a realizar el hecho" (art. 28 par. 2.a CP)
Enunciado BC5 (E_{T5}) = "Persona que coopera a la realización del hecho" (art. 28 par. 2.b CP)

De este modo, el denominado "Base de Conocimiento" estaría formada, en este caso, por el conjunto siguiente: BC = { $E_{BC1} \Lambda E_{BC2} \Lambda E_{BC3} \Lambda E_{BC4} \Lambda E_{BC5}$}. Este conjunto BC obviamente se puede transformar en un conjunto de reglas de producción convirtiendo cada uno de los anteriores enunciados con la estructura "Si _____ entonces _____". De este modo, el conjunto BC estaría formado ahora por un conjunto de reglas de producción susceptibles de ser analizadas por un computador.

A dicho conjunto denominado BC, y específicamente en el ámbito de la aplicación del Derecho, se le tiene que unir otro conjunto que podríamos denominar P formado también por enunciados pero en este caso tendrían naturaleza o contenido empírico recogiendo los hechos que se consideran probados a través de los medios de prueba admitidos en el proceso. Un ejemplo de este conjunto P sería el compuesto por un único enunciado:

Enunciado P1 (E_{P1}) = "x afirmó con otras personas y en su balance que la entidad Y tenía unos beneficios de 3.500 M de euros cuando la realidad es que Y tuvo unas pérdidas de 20.000 M de euros"

Así, el conjunto P, simplificándolo al máximo, se representaría del siguiente modo: P = { E_{P1}}.

Disponiendo ya de los anteriores elementos se pueden presentar un razonamiento abductivo completo en la esfera jurídica (en concreto dentro del ámbito penal). Dicho razonamiento tendría la siguiente estructura:

BC = conjunto de enunciados normativos que definen cuándo una persona es responsable criminalmente de un delito o falta.

P = conjunto de enunciados fácticos que recogen los hechos empíricos probados, es decir, que reflejan una realidad.

H = hipótesis que explica un hecho "sorprendente" necesario de explicación. En este caso dicha hipótesis, teniendo en cuenta las presunciones legales, se partiría de la presunción de inocencia recogida en el art. 24.2 de la Constitución española por lo que dicha hipótesis de partida estaría formada por el enunciado "x es inocente" que sería equivalente a afirmar que dicho x no es responsable criminalmente del delito o falta que se le imputa.

h = hecho "sorprendente" (en terminología de Peirce) que necesita ser explicado. En este caso podría ser el enunciado siguiente: "El banco Y quebró al tener unas pérdidas de más de 20.000 M de euros".

Desde un punto de vista formal dicho argumento se formularía al inicio del juicio de la siguiente manera: $\{BC \wedge P \wedge ?\} \models h$. Lo necesario es formular una hipótesis (abducir) que sea capaz de explicar h teniendo en cuenta BC y P. Por ello, en un segundo momento el anterior esquema argumentativo se completaría con dicha hipótesis adquiriendo la siguiente forma: $\{BC \wedge P \wedge H\} \models h$.

Dado que la hipótesis propuesta (H = "x es inocente") no es consistente con BC y P concretamente con el E_{BC2} es preciso rechazar dicha hipótesis explicativa y afirmar su contraria, es decir, que "x NO es inocente" y por tanto responsable del delito que se le acusa (en este ejemplo el de quebrar el banco maliciosamente al falsear sus balances).

Llegado a este punto, dicho esquema de inferencia puede ser analizada su corrección por cualquier método lógico siendo de especial interés el empleo de algún método de decisión como por ejemplo los árboles semánticos (*tableaux*). Y ello es así dado que un método de decisión permite resolver la corrección o no de un argumento a través de un número finito de pasos y de forma automática, es decir, que sea un algoritmo. Lo anterior tiene la ventaja de permitir que dicho análisis lógico pueda ser analizado computacionalmente.

Si se emplea el método de decisión antes indicado nos podemos encontrar con que todas las ramas cierran en cuyo caso lo que se nos está indicando es que la hipótesis (H) no es consistente dado que nos lleva, en todos los casos, a contradicciones teniendo que rechazarla y aceptar su contraria. Si alguna de las ramas quedasen abiertas ello significaría que en , al menos un caso, dicha hipótesis es consistente por lo que habría que seguir manteniendo la inocencia del acusado.

Las conclusiones a las que pueden llegar es que el razonamiento abductivo responde a un tipo de razonamiento "reales", es decir, que captan la manera de razonar en muchos ámbitos de la realidad humana. Dicho esquema argumentativo permite captar la manera de argumentar en la esfera jurídica dotándola de una racionalidad sólida al poder emplear cualquier método del cálculo

lógico. Y, por último, permite la aplicación de todas las técnicas propias de la IA a la hora de poder implementar dichos análisis en una computadora.

REFERENCIAS

Aliseda, A. 2006, *Abductive Reasoning. Logical Investigations into Discovery and Explanation,* Dordrecht, Springer.
Hanson, N. R. 1961, *Patterns of Scientific Discovery,* Cambridge at The University Press
Harman, G. 1965, "The Inference to the Best Explanation", *Philosophical Review* 74, 88-95
Peirce, C.S. 1958, *Collected Papers of Charles Sanders Peirce,* Volumes 1-6 edited by C. Hartshorne, P. Weiss, Cambridge, Harvard University Press, 1931-1935; and volumes 7 – 8 edited by A.W. Burks, Cambridge, Harvard University Press
Reilly, F. E. 1970, *Charles Peirce's Theory of Scientific Method,* New York, Fordham University Press

GÖDEL'S SYNTACTICAL CONVENTIONS

PIERDANIELE GIARETTA, SILVIA GAIO
University of Padova

ABSTRACTIn his criticism of the syntactical "interpretation" of mathematics, which should have appeared in a volume on Rudolf Carnap, Kurt Gödel employs a rather special notion of convention and claims that syntactical conventions should be admissible; i.e. that empirical facts should not follow from them. Both the notions of convention and of admissibility are tentatively clarified on the basis of what Gödel says. As concerns admissibility, the underlying idea is that conventions are to be void of empirical content. Two alternative readings of voidness of empirical content are isolated. However, for different reasons, both appear to be unsatisfactory.

KEYWORDS
Gödel, syntactical interpretation, convention, empirical content

1. A central point of Gödel's unpublished criticism on the syntactical "interpretation" of mathematics, which should have appeared in a volume on Rudolf Carnap, is the thesis that empirical facts cannot follow from syntactical conventions. Gödel presents this requirement as being involved in the notion of syntactical convention and holds that it implies that "the rules of syntax must be demonstrably consistent, since from an inconsistency *every* proposition follows, all factual propositions included" (Gödel 1995, 339)[1].

Indeed, Gödel's notion of a syntactical convention implies something stronger than consistency; that is *a priori* acknowledgeability of consistency, as is clear from the following passages:

> […] a rule about the truth of sentences can be called *syntactical* only if it is clear from its formulation, or if it somehow can be known beforehand, that it does not imply the truth or falsehood of any "factual" sentence (i.e., one whose truth, owing to the semantical rules of the language, depends on extra-linguistic facts). (339)

> A consistency proof […] is indispensable because it belongs to the concept of a convention that one knows it does not imply any propositions which can be falsified by observation (which, in the case of the mathematical "conventions", is equivalent with consistency […]). (347)

[1] All subsequent quotations are from Gödel, 1995. Only page numbers will be indicated.

Gödel's idea seems to be that one cannot conventionally posit that the mathematical axioms are true unless one knows that no sentence falsifiable by observation follows from them. Gödel says that, in the case of the mathematical "conventions", such knowledge is equivalent to consistency. He is probably thinking of a proof of consistency, not just of consistency.

Gödel's idea needs to be exactly specified. However, even before any exact specification of it, it suggests how it is possible to correct the *non sequitur* contained in Gödel's saying that "the rules of syntax must be demonstrably consistent, since from an inconsistency *every* proposition follows". The *non sequitur* is noticed by Awodey, Carus, 2004, 2010. They correct it by dropping "demonstrably" to thus get: "the rules of syntax *must be consistent*, since from an inconsistency every proposition follows". A different correction, suggested by a broader report of Gödel's conception of a convention, consists in integrating, or making explicit, something Gödel had already said but omitted to say again: the rules of syntax must be demonstrably consistent, since from an inconsistency every proposition follows, all factual propositions included, *and their being conventions implies that it must be evident, or already proved, that they do not imply any factual proposition*.

Previously acknowledged consistency is a problematic point. It might be taken as a necessary, but not sufficient condition for adopting a convention. Such a condition can be stated by means of the following semi-formal principles, where R is a variable ranging over rules (including axioms), C stands for "a convention", K_t for "it is known at t", $K_{\leq t}$ for "it is known at some $t' \leq t$, "cons" for "consistent":

1. R is C at t $\rightarrow K_t$(R is C at t)
2. R is C at t $\rightarrow$ (K_t(R is C at t) $\rightarrow K_{\leq t}$(R is cons))

Principle 1 might be motivated by the consideration that a rule can be a convention at time t only if it is acknowledged as a convention at t. Gödel would probably accept this implication, but someone else, like Carnap, might refrain from accepting it. Being a convention might be conceived as depending only on the way it is formulated and one could adapt to a convention without being aware of the convention or even of the fact that one is adapting to it.

The following is a consequence of 1 and 2, which Gödel appears to endorse explicitly:

3. R is C at t $\rightarrow K_{\leq t}$(R is cons))

Principle 3 says that a rule cannot be a convention without being known, or already known, to be consistent. Principle 3, like 1 and 2, is a bit weird. Principle 3 might appear a bit less weird if one thinks of 3 as somewhat similar to the meta-rule adopted in some States, according to which a statement, or a sequence of statements, is a law only if it has passed a formal control and has been published in the Official Gazette.

Acknowledgement of consistency does not exclude that a rule allowing the statement of false sentences be adopted. An arithmetical translation of the statement that first-order arithmetic – in a suitable formalization– is not consistent might be added to first-order arithmetic. The resulting system allows the derivation of sentences that are false in the standard interpretation, but, as is well known, the system is consistent and could be posited as true.

2. Consistency is a consequence of the underivability of any sentence falsifiable by (empirical) observation, and Gödel takes the latter, i.e., underivability of any sentence falsifiable by (empirical) observation, as the requirement to be satisfied by a convention devised for stating sentences. Such a convention would be expressed by means of syntactical rules and Gödel says they should be admissible in the sense that it must be known that they do not have any empirical consequences:

What must be known is that the rules, by themselves, do not imply the truth or falsehood of any proposition expressing an empirical fact. Such rules may be called "admissible". (357)

The relation between the consistency condition and the requirement of the underivability of any sentence falsifiable by (empirical) observation is not completely clear. Even if the underivability of any sentence falsifiable by (empirical) observation appears to imply consistency (in classical and intuitionistic logic), the way in which Gödel thinks of "propositions which can be falsified by observation" is rather complex.

Gödel does not rely on a neat separation of the mathematical and the empirical fields. Gödel's main point seems to be that, on the one hand, empirical matters, mainly in physical sciences, have some aspects which are intrinsically mathematical and, on the other hand, some basic combinatorics can be interpreted as a theory concerning physical objects. Content that is specifically mathematical would be introduced only where the infinite is involved. To deal with it, mathematical intuition must be resorted to, but applying mathematics might not need mathematical intuition.

Gödel's point does not seem to be that mathematical knowledge does not involve empirical knowledge, but rather that mathematical knowledge, in particular strictly mathematical knowledge, involves non-empirical knowledge. Apparently, for the sake of the argument, but perhaps not only

for that, Gödel seems to grant that mathematics may have an empirical content. What he most strongly criticises is the thesis that mathematics lacks any content at all.

How then is it possible for Gödel to require that no empirical fact be implied by mathematics interpreted as a convention?

In Gödel's view, the conception of mathematics as a syntactic linguistic convention does not avoid implicitly assigning a minimal linguistic empirical content to mathematics. For some empirical knowledge is required to know (and adopt) the rules of syntax ("the objects to which [the finitary rules] refer may be looked upon as parts of the physical world (either existing already or producible)", 338, footnote 14) and such a knowledge is partially expressed by the very rules of syntax. What Gödel seems to require is that nothing empirical beyond that should be implied by those rules. In other words, Gödel is using a relative notion of empirical fact, as is shown by some passages like the following: "Conventions about the use of symbols are void of content only in a *relative* sense, namely, insofar as they *add* nothing to a theory which implies their admissibility" (358); "one can speak of conventions only *relative to a certain system of knowledge*, in the sense that they don't *add* anything to this system" (342). It is clear that voidness of content is relative to what is already known: no further content is implied beyond what is involved by a system of knowledge supporting the conventions.

A further clue for better specifying Gödel's idea is contained in a passage where Gödel refers to the "circumstance" that "pure mathematics implies nothing about the truth values of those sentences which contain no mathematical symbols (i.e., within the field of mathematics and natural science, the atomic empirical sentences)" (352). Following the suggestion given in this passage, we will not speak of empirical facts but of atomic sentences or negations of atomic sentences built from an empirical vocabulary. We assume that an empirical vocabulary contains –besides the identity sign, which is to be included because of the universality of identity– predicates or function expressions standing for properties, relations or functions concerning empirical objects or quantities, and, possibly, expressions referring to empirical objects or quantities.

Saying exactly what makes a vocabulary empirical is not so easy. We might think of its basic expressions under an interpretation directly pointing to some pieces or some aspects of the empirical world. However, especially when science is highly mathematised, an idealised interpretation is given in the first place, and then the empirical world is involved by establishing correspondences, measuring quantities, and determining approximation intervals. We omit dealing with such complexity, as it is not needed for our purposes, and we will speak of an empirical

language by simply intending a language provided with an empirical interpretation in a sense that can be made more or less precise case by case.

3. The above considerations seem to point out a potential conflict between Gödel's admissibility condition about syntactic (and yet mathematical) conventions and his general view of the impossibility of avoiding any content at all. On one side Gödel requires that syntactic (and yet mathematical) conventions "do not imply the truth or falsity of any proposition expressing an empirical fact", that is, no sentence with empirical content follows from syntactical conventions. This condition seems to suggest that nothing empirical can be involved in mathematics. On the other side, he suggests that syntax can be seen as concerning physical objects and lack of content is relative. It is then legitimate to ask: in which sense is mathematics void of empirical content? Moreover, assuming that the sense of empiricalness should be specified case by case with reference to a specific language in some way connected with the physical world, should we say that the language of mathematics cannot be such a language?

If you take for granted that nothing of the mathematical language is related to the physical world, pure mathematics is conservative and does not bring about any new knowledge related to the world; that is, pure mathematics is able to imply neither the truth nor the falsity of any proposition carrying an empirical content. If so, then Gödel's admissibility condition seems to be obvious and unquestionable. In this case, then, admissibility coincides with consistency in the sense that a proof of admissibility is a proof of consistency too and vice versa.

Consider, however, the role played by mathematics in empirical sciences. Mathematical rules applied to certain laws of nature expressed in a mathematical form allow us to obtain other propositions with empirical content. Some of them are also laws of nature and empirical observation is needed to learn about them or to justify them. When mathematics happens to be so integrated into scientific theories about empirical aspects of the world, it turns out to be able to "say" something about the world. Furthermore, Gödel does not seem to exclude an even more direct application of mathematics to the empirical world.

Gödel thinks that the adoption of syntactical conventions involves or presupposes the application, and so the implicit acceptance, of finitary combinatorics: finitary combinatorics seems to be identified with what he calls "the body of knowledge which is indispensable for making any linguistic conventions at all" (342-343). Finitary combinatorics is also a part of mathematics that can be expressed in a formal theory of numbers. Clearly, Gödel takes it as provided with content. When finitary combinatorics is applied to symbols in syntax, Gödel refrains from qualifying its

content as non-empirical or as involving nothing empirical. In general, the languages devised to speak of empirical objects and finite collections, or sequences, of such objects might be seen as extensions of a combinatorial language empirically interpreted. Thus, looking at the finitary combinatorial part of mathematics, its content might be considered as empirical, even if in a somewhat idealised sense, and constitute the basis for the most elementary applications of mathematics to the physical world.

With reference to a specific empirical language L_E, extending a language L of a first order theory T by adding the empirical primitives belonging to some vocabulary V_E, we might consider T as *void of the empirical content of* V_E if and only if, after the rules of T are referred to L_E, no atomic sentences or negations of atomic sentences of L_E, not belonging to the language of finitary combinatorics, are derivable in T from any set of atomic sentences or negations of atomic sentences of L_E, unless they are already derivable from the set by means of pure logic. The core idea is that T is void of the empirical content of V_E if, for any set of empirical data expressed by means of V_E, T does not have the resources to extract from them *new* empirical data not belonging to the set, except those expressible by means of the resources of finitary combinatorics.

The above rough definition is given with reference to a specific empirical content which is supposed to be expressed by a specific empirical language. How is it possible to express in a precise and a coherent way Gödel's general informal thesis that mathematical axioms and rules are admissible in the sense that they "do not imply the truth or falsity of any proposition expressing an empirical fact"?

A possible suggestion seems to come from the following passage already partially quoted:

> If one speaks of conventions and their voidness of content in an absolute sense, this can only mean that they are conventions relative to that body of knowledge which is indispensable for making any linguistic conventions at all. (342-343)

For simplicity we limit ourselves to first order theories including finitary combinatorics. Let L_f be the sublanguage of finite combinatorics of the language L of a theory T. Then we consider all the empirical languages L_E that can be obtained from L_f by adding the empirical primitives belonging to some vocabulary V_E. We might consider T as *void of empirical content* if and only if, after the rules of T are referred to L_E, for every language L_E, as outlined above, no atomic sentence or negation of an atomic sentence of L_E, not built from the vocabulary of L_f, is derivable in T from a set of atomic sentences or negations of atomic sentences of L_E, unless it is already derivable from the set by means of pure logic. In this case, the idea is a generalization of the one underlying the previous rough definition: T is void of empirical content if, for any set of empirical data of any kind, T does

not have the resources to extract from them *new* empirical data; i.e. data not belonging to the set, except those expressible by means of the resources of finitary combinatorics.

Does the given definition provide a satisfactory notion of voidness of empirical content that succeeds in overcoming the limitations of the relative notion? General quantification over languages is clearly a critical point; first of all because of the indeterminacy affecting the general notion of language. In order to reduce indeterminacy, we may quantify only over first order languages provided with the same logical-mathematical vocabulary. Still, it is undetermined what an empirical vocabulary is. Some easily available examples are not enough to specify the meaning of "empirical" in a rigorous way. Even if you take into account only a specific class of languages syntactically identified, the notion of what makes them empirical, i.e. an empirical interpretation, remains indeterminate. So, we cannot say that we really know what it means to quantify over all empirical languages.

The outcome of our analysis is that the relative voidness of empirical content does not really distinguish mathematics from empirical science. At most, it warrants that some empirical facts do not follow from (pure) mathematics. On the other hand, the suggested notion of absolute voidness of empirical content appears to involve a general notion of empirical language which eludes a precise understanding. Moreover, both the relative and the absolute notion presuppose a special status for finitary combinatorics: it is assigned an empirical or quasi-empirical role. Concerning this point, a preliminary satisfactory account is needed.

REFERENCES

Awodey, S., Carus, A. W. 2004, 'How Carnap Could Have Replied to Gödel', in Awodey, S., Klein. C. (eds.), *Carnap Brought Home: the View from Jena*, Chicago, Open Court, 203-223.
Awodey, S., Carus, A. W. 2010, 'Gödel and Carnap', in Feferman K., Parsons, C., Simpson, S.G. (eds), *Kurt Gödel: Essays for His Centennial*, New York, Cambridge University Press, 252-274.
Carnap, R. 1937, *The Logical Syntax of Language*, London, Kegan Paul; English translation of *Logische Syntax der Sprache*, Wien, Verlag von Julius Springer, 1934.
Gödel, K. 1995, *Collected Works*, vol. III, *Unpublished Essays and Lectures*, edited by S. Feferman *et al.*, New York, Oxford University Press.

ABDUCCIÓN Y SEMÁNTICA DE TEORÍA DE JUEGOS

IGNACIO HERNÁNDEZ ANTÓN, ENRIQUE SARRIÓN MORILLO, FERNANDO SOLER TOSCANO
Grupo de Lógica, Lenguaje e Información - U. de Sevilla
iha, esarrion, fsoler@us.es

ABSTRACT: An agency-inspired inferential game is proposed in which two agents (abductive and deductive) interact. It depicts hybrid inferencial contexts where deductive and abductive steps are mixed. The abductive agent can use results obtained by the deductive one but not vice versa due to the abductive process do not always generate deductive valid formulas..

Key words: Abductive reasoning, Deduction, Logic games, Agents.

1 Introducción

En el tratamiento formal del razonamiento abductivo aparece frecuentemente la cuestión de la relación entre deducción y abducción (Soler-Toscano and Nepomuceno-Fernández 2008). Ambas formas de inferencia parecen estar abocadas a convivir juntas, a pesar de que sus propiedades son completamente diferentes. Así, las aproximaciones lógicas a la abducción se basan más o menos en lo que podemos considerar como *usos abductivos de cálculos deductivos* (Soler-Toscano, Nepomuceno-Fernández, and Aliseda-Llera 2009). El planteamiento que defendemos en este trabajo es que el interés no se encuentra tanto en la definición de cálculos abductivos *puros* como en tratar contextos inferenciales *híbridos* en los que se mezclan procesos deductivos y abductivos.

Dado un lenguaje formal $\mathscr{L}$ y una relación de consecuencia lógica $\vdash$, es habitual considerar que (Θ, φ) es un problema abductivo, siendo $\Theta \subset \mathscr{L}, \Theta \nvdash \bot$ y $\varphi \in \mathscr{L}$, si y sólo si[1]

$$\Theta \nvdash \varphi \tag{1}$$

La *solución* a un problema abductivo (Θ, φ) viene dada por una fórmula $\mathscr{L}$ tal que

$$\Theta \cup \{\alpha\} \vdash \varphi \tag{2}$$

[1]En (Aliseda 2006) se distingue entre novedad y anomalía como dos tipos de problemas abductivos. La novedad coincide exactamente con (1) pero en la anomalía se exige, además, $\Theta \vdash \neg\varphi$.

Como se puede observar, las definiciones de "problema abductivo" (1) y "solución abductiva" (2) dependen de la relación de consecuencia $\vdash$ (lo habitual es considerar que esta relación es la de la Lógica Clásica, pero, naturalmente, se pueden pensar en otras posibilidades distintas). Así pues, las propias nociones relativas a la inferencia abductiva se construyen a partir de la inferencia deductiva. Sin embargo, si vemos por ejemplo (Hintikka 1998), uno de los requisitos que debe cumplir la abducción es su autonomía con respecto a la deducción[2]. Por tanto, existe una tensión entre los planteamientos lógicos -que requienen de la deducción para definir la abducción-, y los epistemológicos -que reivindican la autonomía de ambas formas de razonamiento-.

No hay duda de que los contextos inferenciales en los que habitualmente aparece el razonamiento abductivo son lo suficientemente ricos como para mezclar diversos tipos de inferencias. Por ello, el interés no está tanto en los contextos *puros*, donde tal vez se dé la independencia entre ambas formas de razonamiento, sino en los *híbridos*, dada la gran habilidad con que los sujetos somos capaces de combinar procesos de inferencia deductiva y abductiva.

Debemos señalar que, mientras que a las inferencias deductivas se les impone el requisito de la corrección, las abductivas se caracterizan por seguir lo que podríamos llamar la *regla de Peirce*,

$$\frac{\alpha \to \varphi, \qquad \varphi}{\alpha} \tag{3}$$

que es claramente incorrecta desde el punto de vista deductivo[3].

El problema que abordamos en este trabajo es, precisamente, el de la combinación de ambas formas de inferencia. Para ello necesitamos llevar en paralelo dos líneas de razonamiento: la deductiva -que no emplea la *regla de Peirce*- y la abductiva -que la usa y propone sus conclusiones como meras hipótesis-. Ambas discurren a la vez, una produciendo resultados necesarios y la otra conjeturas. Debemos tener en cuenta que no se pueden reutilizar las inferencias obtenidas en la línea abductiva como premisas en la deductiva, debido al carácter hipotético de los resultados inferidos en la primera de las líneas citadas.

Para dividir ambas líneas, e inspirándonos en la Semántica de Teoría de Juegos (Hintikka and Saarinen 1979), consideraremos dos jugadores, a saber: $\mathscr{D}$ (deductivo) y $\mathscr{A}$ (abductivo). Establecemos además que los pasos inferenciales de cada agente están sujetos a *reglas operatorias*. El principal proceso inferencial del jugador abductivo consistirá, justamente, en la aplicación de una variante de la mencionada *regla de Peirce*, tomando para ello como premisas una fórmula propia y otra generada

[2]La abducción es realmente un proceso irreductible a la deducción (e igualmente a la inducción).
[3]Es la conocida falacia de afirmación del consecuente.

por $\mathscr{D}$. De otro lado, el jugador deductivo operará sobre la teoría Θ para producir fórmulas que puedan servir a $\mathscr{A}$ en su búsqueda de explicaciones.

2 El comienzo del juego y sus reglas

El juego es de naturaleza dialógica y comienza con el objetivo general de resolver un problema abductivo (Θ, φ), estableciéndose que $\mathscr{D}$ parte de las fórmulas de Θ y $\mathscr{A}$ de la fórmula φ. En cuanto a las reglas que rigen el diálogo entre los agentes $\mathscr{D}$ y $\mathscr{A}$, consideraremos en este artículo que trabajamos en lógica proposicional clásica, aunque las reglas que mostramos se pueden extender fácilmente a otros sistemas. Además de las *reglas operatorias*, propias de cada conectiva lógica, consideramos *reglas estratégicas* que indican el modo en que se llevan a cabo los diálogos abductivos, incluyendo la forma en que empiezan y acaban éstos.

2.1 Reglas para $\mathscr{D}$

El jugador $\mathscr{D}$ comienza a jugar con las fórmulas de Θ y puede realizar pasos deductivos que van extrayendo fórmulas que son consecuencia lógica de Θ. Para esto puede usar cualquier cálculo deductivo, por lo que no detallamos unas reglas concretas. Una opción que resulta útil en combinación con el proceso abductivo es emplear cálculos basados en resolución tal y como se indica en (Soler-Toscano, Nepomuceno-Fernández, and Aliseda-Llera 2006).

2.2 Reglas para $\mathscr{A}$

El jugador $\mathscr{A}$ parte del hecho que se quiere explicar y juega *por objetivos explicativos*. Usará las fórmulas generadas por $\mathscr{D}$ para producir posibles explicaciones que se irán refinando poco a poco. El principal paso de inferencia de $\mathscr{A}$ es la regla abductiva:

$$\frac{\mathscr{D} : \overline{\alpha} \vee \varphi, \qquad \cancel{\mathscr{A} : \varphi}}{\mathscr{A} : \alpha} \tag{4}$$

donde α y $\overline{\alpha}$ representan, respectivamente, una fórmula y su negación[4]. Como se puede apreciar, la aplicación de (4) se realiza tomando siempre una premisa de $\mathscr{D}$ y otra de $\mathscr{A}$. La conclusión de la regla (4) es también una fórmula de $\mathscr{A}$, ya que no se trata de una consecuencia deductiva de las premisas, sino abductiva. La premisa $\mathscr{A} : \varphi$ que se ha usado para aplicar (4) ya no vuelve a utilizarse, es decir, se ignora tras la aplicación de la regla (ése es el sentido de que dicha fórmula aparezca tachada). Así pues, decimos que φ queda *explicada* por α, por lo que el nuevo objetivo del juego será explicar α.

El jugador $\mathscr{A}$ cuenta también con *reglas de simplificación* que puede aplicar a sus objetivos:

[4]Obsérvese que (4) es equivalente a la regla de Peirce (3).

$$\frac{\cancel{\mathscr{A} : \alpha \wedge \beta}}{\mathscr{A} : \alpha \quad \mathscr{A} : \beta} \tag{5}$$

$$\frac{\cancel{\mathscr{A} : \alpha \vee \beta}}{\mathscr{A} : \alpha} \tag{6}$$

$$\frac{\cancel{\mathscr{A} : \alpha \vee \beta}}{\mathscr{A} : \beta} \tag{7}$$

Téngase en cuenta que, cuando $\mathscr{A}$ tiene que explicar una disyunción, debe elegir entre dos reglas que son distintas en general -la (6) y la (7)-, lo cual conlleva que la búsqueda de soluciones abductivas no sea determinista. Sin embargo, para explicar una conjunción, la regla (5) exige que se expliquen ambos términos.

Adicionalmente damos a $\mathscr{A}$ una regla de confirmación, que deja de buscar explicaciones para un objetivo que ha sido encontrado por $\mathscr{D}$:

$$\frac{\mathscr{D} : \alpha \qquad \cancel{\mathscr{A} : \alpha}}{\mathscr{A} : \bot} \tag{8}$$

Lo que hace la regla (8) es detener el proceso de búsqueda si el objetivo que $\mathscr{A}$ tiene que explicar, α, ha sido deducido por $\mathscr{D}$. Téngase en cuenta que esto no implica que el diálogo termine, porque aún puede haber algún objetivo abierto.

Además, $\mathscr{A}$ puede aplicar *reglas de transformación* a sus objetivos para convertirlos en fórmulas equivalentes:

$$\frac{\cancel{\mathscr{A} : \alpha \rightarrow \beta}}{\mathscr{A} : \neg\alpha \vee \beta} \tag{9}$$

$$\frac{\cancel{\mathscr{A} : \neg(\alpha \rightarrow \beta)}}{\mathscr{A} : \alpha \wedge \neg\beta} \tag{10}$$

$$\frac{\cancel{\mathscr{A} : \neg(\alpha \wedge \beta)}}{\mathscr{A} : \neg\alpha \vee \neg\beta} \tag{11}$$

$$\frac{\cancel{\mathscr{A} : \neg(\alpha \vee \beta)}}{\mathscr{A} : \neg\alpha \wedge \neg\beta} \tag{12}$$

2.3 *Reglas de terminación* del juego

Finalmente, vamos a ver las distintas condiciones de terminación del juego[5]:

[5]No se debe confundir con la *regla de confirmación* de un objetivo (8).

- Si aparecen las fórmulas $\mathscr{D} : \alpha$ y $\mathscr{D} : \overline{a}$, el proceso termina porque la teoría es inconsistente. Para ello, usamos la regla:

$$\frac{\mathscr{D} : \alpha \qquad \mathscr{D} : \overline{\alpha}}{\mathscr{D} : \bot} \tag{13}$$

- Si aparecen $\mathscr{D} : \alpha$ y $\mathscr{A} : \overline{\alpha}$, se termina porque la explicación que se está construyendo es inconsistente con la teoría. Esto lo hace la regla:

$$\frac{\mathscr{D} : \alpha \qquad \mathscr{A} : \overline{\alpha}}{\mathscr{A} : \bot} \tag{14}$$

- Si todas las fórmulas que puede obtener $\mathscr{A}$ son de la forma $\mathscr{A} : \top$, el proceso termina porque la teoría explica la observación y, por tanto, no hay verdadero problema abductivo.

- En otro caso, la conjunción de las fórmulas que quedan en la línea $\mathscr{A}$ puede proponerse como una posible solución abductiva compuesta.

2.4 Ejemplo de juego

La tabla 1 muestra como ejemplo la resolución del problema abductivo $(\{p, p \to q, q \wedge r \to s\}, s)$. Cuando el diálogo concluye, $\mathscr{A}$ tiene r como único objetivo, que puede ser propuesto como solución abductiva.

Tabla 1: ejemplo de resolución de un problema abductivo

DEDUCCIÓN		ABDUCCIÓN	
$\mathscr{D}_1 : p$	(teoría)	$\cancel{\mathscr{A}_1 : s}$	(observación)
$\mathscr{D}_2 : p \to q$	(teoría)	$\cancel{\mathscr{A}_2 : q \wedge r}$	(de $\mathscr{D}_5$ y $\mathscr{A}_1$)
$\mathscr{D}_3 : q \wedge r \to s$	(teoría)	$\cancel{\mathscr{A}_3 : q}$	(de $\mathscr{A}_2$)
$\mathscr{D}_4 : q$	(de $\mathscr{D}_1$ y $\mathscr{D}_2$)	$\mathscr{A}_4 : r$	(de $\mathscr{A}_2$)
$\mathscr{D}_5 : \neg(q \wedge r) \vee s$	(de $\mathscr{D}_3$)	$\mathscr{A}_5 : \top$	(de $\mathscr{D}_4$ y $\mathscr{A}_3$)

3 Tipos de solución abductiva

Es frecuente, dado un problema abductivo (Θ, φ), clasificar las distintas soluciones abductivas posibles y en distintos tipos según satisfagan o no ciertas condiciones:

- **Consistencia.** Se exige la consistencia de ψ con Θ, es decir, $\Theta \cup \{\psi\} \nvDash \bot$. Como hemos visto más arriba, una de las condiciones de terminación del cálculo estriba en que la solución obtenida sea inconsistente con la teoría. Por tanto, los diálogos abductivos tienen la capacidad de detectar soluciones inconsistentes.

- **Explicatividad.** Se trata de evitar soluciones triviales, las cuales explican la observación sin necesidad de la teoría. Para ello, se suele exigir $\psi \not\models \varphi$. Si $\mathscr{A}$ no aplica ninguna regla, entonces la explicación que obtiene es trivial: se trata de la misma observación que se quiere explicar. A medida que $\mathscr{A}$ emplea más reglas, las soluciones que se obtienen son más explicativas, porque usan más fórmulas de entre las generadas por $\mathscr{D}$.

No hemos abordado aún el problema de la *selección de la mejor explicación*, que suele ser uno de los mayores escollos en el tratamiento lógico del razonamiento abductivo. Como hemos tenido ocasión de comprobar, la solución abductiva que se obtiene depende de las decisiones que $\mathscr{D}$ y $\mathscr{A}$ vayan tomando durante el juego -particularmente de la fórmula de $\mathscr{D}$ que $\mathscr{A}$ elija como premisa de (4) (en caso de haber varias) y de la decisión que $\mathscr{A}$ tome, ante una disyunción, de usar (6) o (7)-. Si pensamos en $\mathscr{A}$ como un agente humano, nuestro planteamiento no se diferencia mucho de la forma en que los sujetos generamos y seleccionamos explicaciones. Por lo tanto, la cuestión de "la mejor explicación" queda circunscrita a la determinación de qué estrategias se usan para la ejecución del juego.

4 Observaciones finales

Hemos presentado una primera aproximación al tratamiento dialógico de la tarea abductiva en contextos *híbridos*, en los cuales están presentes inferencias que pueden ser tanto abductivas como deductivas. Estos contextos son más naturales y presentan una mayor semajanza con los contextos científicos, puesto que en estos últimos los procesos inferenciales admitidos incluyen cuando menos los deductivos, abductivos e inductivos. En nuestra propuesta, las tareas inferenciales son realizadas por dos agentes cuyas *reglas estratégicas* aún quedan pendientes para futuros trabajos, pero cuyas *reglas operacionales* quedan bien representadas por las descritas en el presente. Así pues, con el planteamiento descrito se consigue aunar dos tipos de procesos inferenciales que habitualmente tienen un tratamiento casi completamente separado, tanto en el ámbito de la Lógica como en el de la Epistemología.

Agradecimientos

Este trabajo se ha realizado en el marco de ejecución de los proyectos *Interpretaciones Alternativas de Lógicas no Clásicas* (referencia HUM5844 de la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía), y *Conciencia, Lógica y Computación* (referencia FFI2011-29609-C02-01, del Ministerio de Ciencia e Innovación).

REFERENCIAS

Atocha, A. (2006). *Abductive Reasoning: Logical Investigations into Discovery and Explanation*. Springer. Volume 330 of Synthese Lybrary.

Hintikka, J. (1998). What is abduction? the fundamental problem of contemporary epistemology. *Transactions of the Charles S. Peirce Society*, *34*(3), 503–533.

Hintikka, J. & Saarinen, E. (1979). *Game-theoretical semantics: essays on semantics*. D. Reidel Pub. Co. Synthese language library.

Soler-Toscano, F. & Ángel Nepomuceno-Fernández (2008). Deducción y abducción. *Teorema*, *27*(1), 5–16.

Soler-Toscano, F., Ángel Nepomuceno-Fernández, & Aliseda-Llera, A. (2006). Model-based abduction via dual resolution. *Logic Journal of IGPL*, *14*(2), –.

Soler-Toscano, F., Ángel Nepomuceno-Fernández, & Aliseda-Llera, A. (2009). Abduction via c-tableaux and δ-resolution. *Journal of Applied Non-Classical Logics*, *19*(2), 211–225.

FIRST-PERSON LOGICAL THEORIES AND THIRD-PERSON THEORIES

JOHN T. KEARNS
U. at Buffalo

ABSTRACT: A first-person account, a description, explanation, or theory, is an account of experience or actions from the perspective of the person having the experience or performing the actions. A first-person account is really for that first person, and can only be evaluated from the perspective of the person whose account it is. A third-person account is the same for everyone, and can be evaluated from a neutral, or impersonal standpoint. Certain ethical theories and logical theories have an essentially first-person character. This does not make them capricious or subjective, for a first-person theory can be governed by objective principles which are the same for everyone. The theories of illocutionary logic, the logic of speech acts, are essentially first-person theories. These theories explore, both semantically and deductively, the commitments generated by performing assertive illocutionary acts. The present paper argues that even standard logical theories are first-person theories. They have, superficially, the appearance of third-person theories, but they are actually truncated and abstract versions of explicitly first-person theories.

Key words: Illocutionary logic, Philosophical logic, Speech acts.

1 Two Kinds of Theory

We can distinguish two kinds of description, or explanation, or theory of something. Let us call descriptions, explanations, and theories *accounts*. A third-person account is the same for all people who know or accept the account, and even for all people who know about the account. The Ptolemaic theory of the Earth, Sun, planets and fixed stars is a third-person theory. It says how the various heavenly bodies are located, and how they move with respect to one another. The theory doesn't reflect any particular person's perspective.

A *first-person account* of something reflects or incorporates a particular person's perspective on that something. It is an account which only that person could give or adopt, or even consider giving. The account treats something that only the particular person can experience or access, or that only she can access in a way that is relevant to the account. Each person has her own experience. Each person's account of her own perceptual experience, from the "inside" as it were, is essentially first-person. Such an account might be concerned with the objects perceived and their perceived properties and relations, with interpreting her perceptions, or with how she does or can direct her attention to various objects, events, and features. However, while each person's account is essentially

her own, certain principles may govern, or apply to, or be true of everyone's first-person account. For everyone's experience, for example, nothing is or can be red all over and green all over.

Just as no one can have, or share, someone else's experience, so each person's actions are her own. One person can act on behalf of another, but no person can perform another person's actions. There might be a third-person account of people and their behaviors "from the outside," but each person has her own perspective on her own acts, and this enables her to give a first-person account of her own acts and actions. One couldn't even give a third-person account of acts and actions, if it were concerned with more than observable behavior, without borrowing elements from her own first-person understanding.

A first-person account of one's own actions might be historical in the sense that it is concerned with what one has done and tried to do, and with her successes and failures. A first-person theory of action might be normative, incorporating principles about what are the right, or correct, things to do, and with how one must, or ought to, carry out certain acts and activities. A normative theory of this sort can incorporate principles which are the same for everyone, but indicate how the first person involved must, or should, implement the principles in her own particular circumstances. In (Smith 2003), Michael Smith has explained how an essentially first-person ethical theory might incorporate some evaluative principles that each person will use to determine the same positive and negative evaluations for a certain range of situations. The first-person theory will incorporate a sub-theory which has a somewhat third-person "look," because in that sub-theory the evaluations won't differ from one person to the next.

2 Logical Theories

Logical theories can also be classified as first-person theories or third-person theories, although I will argue that logical theories are essentially first-person theories, but often incorporate sub-theories that have a somewhat third-person character. I understand a logical theory to consist of a specialized formal language, usually artificial, a semantic account for that language, and a deductive system for constructing arguments with expressions of the formal language.

Illocutionary logic, the logic of speech acts, or language acts, is explicitly and evidently a first-person theory. Language acts are meaningful acts performed by using expressions, whether the expressions are spoken out loud, written, or merely thought. Both the person who produces the expressions and the person who reads written expressions or who listens with understanding to spoken expressions are performing language acts, but we shall adopt the perspective of the person who produces the expressions that are used. Every language act is some particular person's language act,

but some types of language act –for example, an act of predicating a certain expression of an object– are essentially the same when performed by different people. In studying these language acts, we can ignore the matter of who it is that performs them. Other language acts, such as the illocutionary acts assertion and denial, have consequences for the person who performs them, which consequences do not ”apply” to anyone else. A theory of these acts will record the assertions and denials of a particular language user, and uncover principles she can use to explore the consequences of her own acts.

Logical theories are particularly concerned with *sentential acts*, acts performed with sentences or sentential clauses, and among sentential acts, the theories are concerned with those acts that can appropriately be evaluated in terms of truth and falsity. I shall understand a *statement* to be a sentential act that can appropriately be evaluated in terms of truth and falsity, so that in my sense, to say that someone has made a certain statement is *not* to say that the someone accepts the statement that she made. In my sense, a person can make a statement in order to consider, or even in order to reject, that statement.

A sentential act can be performed with this or that force, like the force of a promise, an apology, or an assertion, to constitute an *illocutionary act*. This is something like the complete act that a person intends to be performing when she performs a sentential act. Statements can also be performed with one or another illocutionary force. For example, a statement can be accepted, or asserted, it can be denied, and it can be supposed to be or not to be the case. An *assertive* illocutionary act is one performed with a statement or statements, and is concerned with the issue of whether the statement(s) fit the world.

Standard logical theories can (and should) be understood to be theories of language acts. But these theories focus on acts which are essentially the same when performed by different people, and end up by abstracting away from the people performing the acts. Such theories generally focus on statements, on the truth conditions of statements, and on semantic features which are determined by, or can be defined in terms of, truth conditions. These are features like implication, logical consequence, and inconsistency.

By studying statements and their features, but ignoring the people who make the statements, logicians produce theories which have a somewhat third-person character. And this has led philosophers and logicians to invent, or postulate –to *fantasize*, really– a realm of abstract propositions which are the eternal bearers of truth and falsity, and are merely “grasped” by ordinary mortals like ourselves. For if logical theories are to be genuinely third-person theories, then they must have objective and impersonal objects to theorize about. Propositions, whether they are the kind that Frege dreamed about or the kind sometimes attributed to Russell, are thought to have the objective character possessed by

molecules, atoms, and particles, though theirs is a somewhat more splendid nature.

3 Illocutionary Theories

A standard logical language contains non-logical expressions, some distinctively logical expressions like connectives and quantifiers, and sentences composed from them. The language is "interpreted" by functions with respect to a domain of objects or with respect to a whole universe of worlds. And there is a deductive system for tracing or tracking logical consequence. A fully worked-out theory of illocutionary logic is obtained from a standard theory by adding things to it.

Sentences in the standard theory's logical language represent statements. The standard theory's language is enriched with various first-person expressions to obtain the language of the full illocutionary theory. Most prominent, perhaps, are the *illocutionary operators* with which a speaker makes explicit the force of her own assertive acts. These are prefixed to the (*plain*) sentences which represent statements to obtain *completed sentences*, which represent illocutionary acts being performed by the speaker. The basic illocutionary operators are the following:

$\vdash$ — the sign of assertion $\llcorner$ — the sign of positive supposition

$\dashv$ — the sign of denial $\neg$ — the sign of negative supposition

An *assertion* is understood to be an act of producing a statement and accepting that statement as being or representing what is the case, or else an act of producing a statement and reaffirming one's continued acceptance of that statement. On this conception, while an assertion can have an addressee, it doesn't require an addressee. A denial is an act of *ruling out* the acceptance of a statement, because the statement fails to fit the world. Suppositions are acts of temporarily accepting or rejecting statements. If a conclusion is deduced from premisses which include suppositions, that conclusion is also a supposition.

Genuine arguments, arguments in real life as opposed to those in logic books and logic courses, are *speech act arguments*. The premisses are illocutionary acts and so are the conclusions. The deductive system in a fully developed illocutionary theory enables a person to construct and study speech act arguments, but she is studying these arguments from her own perspective. They are arguments from her own assertions, denials, and suppositions to conclusions which she herself asserts, denies, or supposes.

4 Illocutionary Semantics

Statements are true or false (and sometimes, perhaps, neither one), statements imply and are incompatible with one another. Assertive illocutionary acts are not true or false, although their component

statements are (usually) one or the other, and assertive illocutionary acts are not linked by implication or incompatibility. But they do possess features which are relevant for evaluating arguments in which they occur. The fundamental feature of this type is *rational commitment*.

Rational commitment is a commitment to do or not do something, or to continue in a certain state like that of accepting a given statement. Deciding to perform an action commits a person to perform that action. Some commitments have a moral dimension, but not all commitments are moral commitments. Some commitments are conditional while others are not. My commitment to close the upstairs windows if it rains while I am at home is conditional, but decisions often establish unconditional commitments.

Performing some acts often commits a person to perform others. Accepting or asserting a given statement will commit a person to accept other statements. These commitments are conditional on the person giving the matter some thought, and having an interest in them. Asserting that today is Wednesday commits me to accept the statement that either today is Wednesday or it is now snowing in Beijing. But I have no interest in that further statement, and would not normally consider it. Still, it would be irrational to accept that today is Wednesday and refuse to grant that either today is Wednesday or it is now snowing in Beijing.

At a given time, a particular person will be committed by her knowledge and belief, including her knowledge of the language, to assert some statements and deny others. With respect to a person at a time, the plus sign '+' is used for the value *committed to perform*. *Commitment valuations* assign this value to completed sentences of the illocutionary language in ways we can specify, and this enables us to determine what it takes for a speech act argument to be *deductively correct*, and further enables us to design deductive systems which can be shown to be sound and complete.

Because logic is a normative subject, the semantic account is provided for an idealized language user whom I call the *designated subject*, for whom I use feminine pronouns. The account designed for this person allows us to determine which arguments are correct for everyone who has the relevant beliefs and disbeliefs, and who (perfectly) understands the language she is speaking.

There is a sense in which each person has her own illocutionary logical theory. This is the theory which incorporates or accommodates her own knowledge and belief at a given time. Although hers is evidently and explicitly a first-person logical theory, each of us can determine which arguments are correct from her perspective. For the deductive principles governing speech-act arguments are the same for everyone, Even standard logical theories are first-person theories, for they are concerned with speech acts, which are always performed by particular people. In standard theories, we can safely ignore which people these are, but we can't get them out of the picture. And standard theories

are concerned with truth-conditional features which don't differ from one time to another.

5 Some Other First-Person Operators

There are a variety of operators which are significant for logic, and which cannot be adequately accommodated in standard logical theories. In (Kearns 2006), I have developed a system for conditional assertions, which constitute a distinctive form of illocutionary acts. Many unsuccessful accounts of ordinary-language conditionals construe these as a special type of statement, rather than as a kind of illocutionary act. If ordinary conditionals were statements, then a given conditional statement would be true or false, no matter who makes it. But a conditional assertion can be in "in effect" for one person and not another, and its being correct or justified for one person makes no demands on anyone else.

Epistemic modal logic is an essentially first-person subject matter, and suitable theories are essentially first-person theories. On no reasonable construal of knowledge or belief does a person know the consequences of everything she knows, or believe the consequences of everything she believes. But a person is committed by what she knows and believes to accept many statements and to deny many others. If '$\Box^K A$' has the significance *I am committed by my current knowledge to accept* A, and '$\Box^B A$' has the significance *I am committed by my current (justified) beliefs to accept* A, then the following inference principles are evidently correct when, first, the premiss assertion has the force of a knowledge claim, and, second, the premiss assertion has the force of a claim to justifiably believe A:

$$\frac{\vdash A}{\vdash \Box^K A} \qquad \frac{\vdash A}{\vdash \Box^B A}$$

With this understanding, the following are also evidently correct:

$$\frac{\vdash \Box^K A \qquad \vdash \Box^K [A \supset B]}{\vdash \Box^K B} \qquad \frac{\vdash \Box^B A \qquad \vdash \Box^B [A \supset B]}{\vdash \Box^B B}$$

A properly conceived theory of epistemic modal logic is a first-person account enabling a person to understand and explore the consequences of her own knowledge and belief.

To provide semantic accounts for first-person theories, we employ functions which are "tailored" to this or that person, characterizing, say, the assertions and denials that a person is committed to perform on the basis of her current knowledge or belief. These semantic accounts, and functions, are *momentary* in the sense that they depend on the subject's knowledge or belief at one or another particular time.

We use language to perform illocutionary acts, and assertive illocutionary acts are linked to one another by rational commitment. This is explored in the explicitly first-person theories of illocutionary logic. We also use language to direct our attention to particular objects and events. A person can *directly* attend to an object which she is presently perceiving, and she can exploit a causal connection that she knows about to directly attend to an object to which that connection links her, which object need not be present. A *referring act* is one which uses an expression to attend in this way. A person can *indirectly* attend to an object by representing certain features and using these features to focus on whatever object possesses them.

A descriptive expression can be used to both represent and refer to an object all at once. It is quite common that the representational component of a description is insufficient to single out the referent ("the chair," "the door"), and a referring expression can even misrepresent the referent. Some referring expressions have no representational content, while some descriptive expressions enable us to indirectly attend to objects ("the tallest woman on Earth") without at the same time directly attending.

A first-person logical theory is needed to accommodate, and account for, the referring use of singular terms. For it is characteristic of the referring use that the speaker exploits a causal connection in the world linking her to the referent. One person can (on different occasions) exploit different connections linking her to a single object, but two people cannot exploit the same connection in referring to an object. An adequate theory of referring must employ different "reference determining" functions for different people.

A full theory of illocutionary logic will be concerned both with assertive illocutionary acts and with reference, and have an explicitly first-person character. A full theory contains a proper subtheory which deals with statements and their truth-conditional semantic features. This subtheory abstracts away from the people who make the statements that are studied, giving the subtheory a third-person "look" or character. But the third-person theory is merely an apparently third-person theory.

REFERENCES

Kearns, J. T. (2006). Conditional assertion, denial, and supposition as illocutionary acts. *Linguistics and Philosophy*, *29*, 455–485.

Smith, M. (2003). Neutral and relative value after moore. *Ethics*, *113*, 576–598.

INCOMPLETENESS, UNDECIDABILITY AND PLATONISM IN THE WORK OF KURT GÖDEL

JOAN ROSSELLÓ

ABSTRACT: This article analyses the relationship between some mathematical and philosophical implications of Gödel's incompleteness theorems, particularly the phenomenon of the inexhaustibility of mathematics and the possible existence of absolutely undecidable sentences, and his Platonism or conceptual realism. In this regard, we conclude that Gödel's proposal to extend Zermelo-Fraenkel's axiomatic set theory with the so-called axioms of infinity, through which he sought to prove hitherto undecidable sentences in set theory and number theory, is closely related not only to his understanding of the phenomenon of incompleteness, but also with his Platonic view on set theory and mathematics in general. The article also aims to describe the development of Gödel's philosophical ideas about the nature of mathematics and their relationship with other foundational programs or philosophical approaches to mathematics such as Carnap's syntactical program, Hilbert's finitism or Husserl's phenomenology. Regarding this, we conclude that Gödel's late interest in Husserl's transcendental idealism responds to the fact that it offers a general philosophical framework for some of Gödel ideas in the philosophy of mathematics and a method for the search and justification of new axioms in mathematics.

KEY WORDS: Incompleteness theorems, inexhaustibility, absolutely undecidable sentences, Platonism, empiricism, finitism, phenomenology, set theory, axioms of infinity

1. Incompleteness theorems and absolutely undecidable sentences

It is well known that Gödel demonstrated in (1931) the incompleteness of every consistent and sufficiently strong theory T (such as PA or ZFC) and the unprovability in such a theory of the statement formalizing "T is consistent." But how did he understand around this time the phenomena of incompleteness? In the footnote 48ª of (1931) (CW I, 181) and several letters and lectures written around the same time, Gödel explains that the addition of an infinite type ω (corresponding to the class of all classes of finite type) to the Russellian theory of types would provide with a definition of truth for the theory investigated and, therefore, prove hitherto undecidable sentences such as those to which refer his incompleteness theorems.

Shortly after Gödel had published his stunning results about the incompleteness of arithmetic, he wrote that although the arithmetical sentence furnished by his theorem is undecidable in the system considered, it is, however, "not at all absolutely undecidable; rather, one can always

pass to "higher" systems in which the sentence in question is decidable. (Some other sentences, of course, nevertheless remain undecidable.)" (*1931?) (CW III, 35). The problem is now that the expressive power of such systems raises the possibility of sentences which "remain undecidable" at *any* level of the hierarchy of the systems considered (think, for example, of the levels of the cumulative hierarchy of sets defined as usual). These are the "absolutely undecidable" sentences referred to in the text above.

There are some well-known candidates for these absolutely undecidable sentences, such as Cantor's continuum hypothesis (CH) or the hypothesis that every set is constructible (for short). Gödel himself considered in (*1939b) (CW III, 155) and other writings of the same period that the statement was absolutely undecidable, but in (1946) (CW II, 151) he came to think that there might be no absolutely undecidable sentences and, therefore, that was surely not absolutely undecidable.

At around the same time, Gödel also conjectured in (1947) that CH was not only consistent with ZFC, as he had previously demonstrated in (1939a), but actually independent, a conjecture that was confirmed by Paul Cohen in 1963. In any case "this would (…) by no means settle the question definitively," but would only show the weakness of the axioms of ZFC in order "to describe some well-determined [set-theoretic] reality" in which CH "must be either true or false" (1947) (CW II, 181).

Therefore, according to Gödel, in order to settle these undecidable sentences in ZFC, such as CH or , we should extend ZFC with new axioms and more precisely with the so called large-cardinal or infinity axioms. According to him, the acceptance of new axioms of infinity could be justified by *intrinsic reasons* (in case they were implied by "the concepts underlying logic and mathematics") or *extrinsic reasons* (in case they had "so abundant verifiable consequences" that they would have to be accepted "quite irrespective of their intrinsic necessity") (Ibid., 182-183).

2. Platonism, incompleteness and undecidability

The Platonic or realist standpoint towards the objects of set theory espoused by Gödel in (1947) had already been defended by him in (1944) to justify his opposition to the constructivist approach taken by Bertrand Russell in his ramified theory of types. Gödel's Platonism in (1944) and (1947) was reinforced in (1964), a revised version of the last article. What is most remarkable of (1964) is the role reserved by Gödel to mathematical intuition in connection with his Platonism. Thus, for example, he argues that mathematical intuition is for mathematicians a source of knowledge as safe as sense perception is for physicists. Actually, Gödel's considered mathematical intuition a primary source of knowledge in the theory of transfinite sets, since it is precisely

intuition what gives rise to the extension of the axioms of ZFC with new axioms of infinity and thus ensures that sentences like CH have meaning.

So Gödel's proposal to extend ZFC with new axioms of infinity was closely related not only to his understanding of the phenomenon of incompleteness, but also to his Platonic view on set theory and mathematics. Gödel's Platonism fits well with the possibility of an intrinsic justification of some axioms of infinity and the meaningfulness of some set-theoretic questions such as CH, but seems completely alien to the possibility of an extrinsic justification of some other.

At the conference (*1951), Gödel offered the most systematic defense of Platonism. He raised again the existence of absolutely undecidable sentences as a possible consequence of his incompleteness theorems and stated that their existence necessarily involves the adoption of a Platonic or realistic point of view regarding mathematical objects. This obviously supposes a stronger notion of absolutely undecidability, since in (1947)/(1964) he had suggested that Platonism implies the decidability and meaningfulness of every hitherto undecidable sentences such as CH and . In (*1951), but also in (*1953/9), Gödel also defended realism against the nominalism or conventionalism of some members of the Vienna Circle, particularly R. Carnap.

Gödel's argues in (*1953/9) that Carnap's view can be refuted by applying the second incompleteness theorem: If the truths of mathematics are to be based solely on syntactical conventions, then these conventions should be consistent (since if this were not the case, all statements would follow from them). Now, by the second incompleteness theorem, a consistency proof for any such set of syntactical conventions would require objects, concepts and methods that go beyond the system in consideration. But now the situation is reversed, since "instead of clarifying the meanings of the non-finitary mathematical terms by explaining them in terms of syntactical rules, non-finitary terms are (used) in order to formulate the syntactical rules" (*1953/9, CW III, 341-442).

3. Empiricism, Finitism and Phenomenology

In the paper on Carnap (*1953/9, footnote 19, CW III, 341), Gödel says that if the syntactical program of Carnap has to be fulfilled, then we must consider it as equivalent to Hilbert's finitist program. In the paper (*1961/?) Gödel criticizes Hilbert's finitist program and sets up Husserl's transcendental philosophy as an alternative to Hilbert's finitist standpoint (and hence to Carnap's syntactical program). He sets up a schema of philosophical views according to their degree and manner of affinity with metaphysics, obtaining a division into two groups, with scepticism, materialism and positivism on the left side, and spiritualism, idealism and theology on

the right side.

According to Gödel, "mathematics, by its nature as an *a priori* science, always has, in and of itself, an inclination toward the right, and, for this reason, has long withstood the spirit of the time [*Zeitgeist*] that has ruled since the Renaissance; i.e., the empiricist theory of mathematics" (*1961/?, CW III, 377). Now, as a reaction to the nihilistic consequences of the prevailing *Zeitgeist*, Gödel continues, came into being "that curious hermaphrodite thing that Hilbert's formalism represents, which sought to do justice both to the spirit of the time and to the nature of mathematics"(Ibid., 379). In Hilbert's finitist program, the rightward, rationalistic beliefs, that proofs provide a secure grounding for propositions and that every clearly formulated mathematical problem is decidable, are combined with the leftward, empirical requisite, that these demands are to be projected into a purely formal system.

However, the first incompleteness theorem shows that for any formal system containing a minimum of arithmetic, there is always an undecidable formula expressible in that system. Moreover, in order that proofs in these formal systems secure the correctness of propositions, they must be consistent, but the second incompleteness theorem shows that to prove the consistency of these formal systems –and hence to provide a secure grounding of propositions– we have to step outside these systems. It turns out that, against Hilbert's hopes, "the certainty of mathematics is to be secured not by proving certain properties by a projection onto material systems –namely, the manipulation of physical symbols but rather by cultivating (deepening) knowledge of the abstract concepts themselves which lead to the setting up of these mechanical systems" (Ibid., 383).

How is it possible, however, to extend our knowledge of these abstract concepts, namely, to make these concepts themselves precise and to gain comprehensive and secure insight into the fundamental relations that subsist among them, i.e., into the axioms that hold for them? We cannot do this by trying to give explicit definitions for concepts and proofs for axioms, since in this case we would need other undefinable abstract concepts and axioms holding for them. Rather, says Gödel, "the procedure must thus consist, at least to a large extent, in a clarification of meaning that does not consist in giving definitions" (Ibid.).

It is precisely at this point that we can appreciate the importance of Husserl's phenomenology as a general philosophical framework for Gödel's ideas on the foundations of mathematics. For, as Gödel acknowledges, Husserl's phenomenology "claims to possess a systematic method for such a clarification of meaning," where "clarification of meaning consists in focusing more sharply on the concepts concerned by directing our attention in a certain way,

namely, onto our own acts in the use of these concepts, onto our powers in carrying out our acts, etc." (Ibid.).

4. Husserl's Phenomenology and Constituted Platonism

We have seen in the previous sections that in different writings from the forties and the fifties, Gödel argues against constructivism, conventionalism, nominalism, and so on, and in favour of Platonism in logic and mathematics. From 1959 onwards Gödel turned his attention to Husserl's transcendental phenomenology. Actually, as Hao Wang suggests, "it seems likely that, in the process of working on his Carnap paper in the 1950s, Gödel had realized that his realism about the conceptual world called for a more solid foundation than he then possessed. At this juncture it was not surprising for him to turn to Husserl's phenomenology, which promises a general framework for justifying certain fundamental beliefs that Gödel shared: realism about the conceptual world, the analogy of concepts and mathematical objects to physical objects, the possibility and importance of categorical intuitions or immediate conceptual knowledge, and the one-sidedness of what Husserl call "the naïve or natural standpoint" (1996, 164-65).

Gödel's professed admiration for Husserl's philosophy raises some issues about the relationship between Gödel's Platonism and Husserl's transcendental phenomenological idealism that ought to be dealt with. Perhaps the main question to be answered is whether Gödel's realism is, after all, in some way compatible with Husserl's transcendental idealism. Richard Tieszen has argued in different places that Gödel's realism about mathematical objects and concepts should be understood not as naïve Platonism, since in this case it would be incompatible with Husserl's transcendental idealism, but as constituted Platonism. Roughly speaking, constituted Platonism asserts that monads (transcendental subjects) constitute in mathematical practice in a motivated and non-arbitrary manner the meaning of being of mathematical objects as mind-independent and abstract. Now, according to Tieszen, "mathematical or logical platonism, in this sense, is compatible with transcendental phenomenological idealism" (2011, 155).

It is difficult to know to what extent Gödel would have been committed to the kind of Platonism suggested by Tieszen. However, given his positive comments on Husserl's philosophy, it seems quite plausible that "Gödel embraced constituted platonism, or was at least moving toward it" (Ibid., 105). Actually, as we have already explained, Gödel's Platonism or realism about the abstract objects and concepts of mathematics was closely linked with the fact that we know about such objects and concepts through mathematical intuition. Gödel often leaves unclear which are the objects of mathematical intuition, but in different places he identifies them with concepts (*1951,

CW III, 30) or concepts and sets (1964, CW II, 271). Now, there are good reasons to think that Gödel's concepts correspond to Husserl's essences (Wang 1996, 167) and even that Gödel understood mathematical intuition in terms of, or in a sense very close to, Husserl's *Wesensanschauung* [categorial or eidetic intuition] (see Ibid., 167-168). Indeed, as we have already noted, it is precisely through mathematical intuition that the clarification of meaning of the concepts we are directing our attention to is given. So it is just as a by-product of this intuition that mathematical objects –number, sets, functions, and so on– are *constituted* by the subject, not as arbitrary free creations of our mind, but as invariant, mind-independent, abstract objects.

5. Concluding remarks: Set Theory as a paradigmatic case

Gödel's Platonism and his phenomenological turn were strongly motivated by the existence of open problems in mathematics, particularly in set theory. For instance, in (*1961/?) Gödel says that there are examples where considerable further development takes place when the phenomenological approach is followed, even without the application of a systematic and conscious procedure. Gödel's example is here the systematic establishment of the axioms of mathematics, where "new axioms, which do not follow by formal logic from those previously established, again and again become evident" (*1961/?, CW III, 385). It is not at all excluded, continues Gödel, that the addition of new axioms that are logically independent from the earlier ones can solve every clearly posed question, since "it is just this becoming evident of more and more new axioms on the basis of the meaning of the primitive notions that a machine cannot imitate" (Ibid.).

It is well known that Gödel asserted in (1947) that reflections on the (iterative) concept of set "of a more profound nature than mathematics is used to giving" are required in order to solve hitherto undecidable sentences such as CH. Now, was Gödel right in thinking that a more profound analysis of the concept of set, following perhaps the lines of Husserl phenomenology, could yield intrinsically justified principles (perhaps other than reflection principles) or axioms that lead to a significant reduction of the class of undecidable sentences? This is still an open question, but some recent results seem to indicate that, if we only admit intrinsic justifications of the axioms and we think these are exhausted by reflection principles, then CH is surely absolutely undecidable (see Koellner 2009). Moreover, there is no intrinsic justification of large cardinal axioms beyond Mahlo's (see Koellner 2006, 162-166) in the sense that they don't seem to be implied in an essential way by the concept of set and so they don't force themselves upon us. The problem is perhaps that our clarification of meaning of the (iterative) concept of set is still not as precise and profound as we need to solve such open problems as CH or, even worse, that the iterative concept of set and CH itself are inherently vague (see, e.g., Feferman 2000, 405).

REFERENCES

Feferman, S. 2000, "Why the programs for new axioms need to be questioned", *Bulletin of Symbolic Logic* 6, 401–413.

Gödel, K. 1929, "Über die Vollständigkeit des Logikkalküls" Dissertation, University of Vienna. In Gödel 1986, 60-101.

—. 1930. "Die Vollständigkeit der Axiome des logischen Funktionenkalküls." *Monatshefte für Mathematik und Physik* 37: 349-360. In Gödel 1986, 102-123.

—. 1931. "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I," *Monatshefte für Mathematik und Physik* 38: 173–198. In Gödel 1986, 144–195.

—. *1931?: "Über unentscheidbare Sätze." In Gödel 1995, 31-35.

—. *1933o. "The present situation in the foundations of mathematics." In Gödel 1995, 45-53.

—. 1938. "The consistency of the axiom of choice and of the generalized continuum hypothesis." *Proceedings of the National Academy of Sciences of the U.S.A.* 24: 556-557. In Gödel 1990, 26-27.

—. 1939a. "Consistency proof for the generalized continuum-hypothesis." *Proceedings of the National Academy of Sciences of the U.S.A.* 25: 220-224. In Gödel 1990, 28-32.

—. *1939b: "Vortrag Göttingen." In Gödel 1995, 127-155.

—. 1940. "The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory." *Annals of mathematics studies*. Vol. 3. Princeton, NJ: Princeton University Press. In Gödel 1990, 33-101.

—. 1944. "Russell's mathematical logic." In Gödel 1990, 119-141.

—. 1946. "Remarks before the Princeton bicentennial conference on problems in mathematics." In Gödel 1990, 150-153.

—. 1947. "What is Cantor's continuum problem?" *The American Mathematical Monthly* 54, 515-527. In Gödel 1990, 176-187.

—. *1951. "Some basic theorems on the foundations of mathematics and their implications." In Gödel 1995, 304-323.

—. *1953/59, III and V. "Is mathematics syntax of language?". In Gödel 1995, 334-363.

—. *1961/?. "The Modern Development of the foundations of mathematics in the light of philosophy." In Gödel 1995, 374-387.

—. 1964. "What is Cantor's continuum problem?" Revised version of Gödel 1947. In Gödel 1990, 254–270.

—. 1986. *Collected Works*. Vol. 1, *Publications 1929–1936*. Edited by S. Feferman et al. New York: Oxford University Press.

—. 1990. *Collected Works*. Vol. 2, *Publications 1938-1974*. Edited by S. Feferman et al. New York: Oxford University Press.

—. 1995. *Collected Works*. Vol. 3, *Unpublished Essays and Lectures*. Edited by S. Feferman et al. New York: Oxford University Press.

Koellner, Peter. 2006. "On the Question of Absolute Undecidability." *Philosophia Mathematica* 14, 2, 153–188.

—. 2009. "On Reflection Principles." *Annals of Pure and Applied Logic* 157, 2-3, 206-219.

—. 2011. *After Gödel. Platonism and Rationalism in Mathematics and Logic*. Oxford, New York: Oxford University Press.

Wang, Hao. 1996. *A Logical Journey: From Gödel to Philosophy*. Cambridge, MA: MIT press.

TRATAMIENTO MULTIMODAL DE CONTEXTOS

ENRIQUE SARRIÓN MORILLO, IGNACIO HERNÁNDEZ ANTÓN, ÁNGEL NEPOMUCENO FERNÁNDEZ
Grupo de Lógica, Lenguaje e Información - U. de Sevilla
esarrion, iha, nepomuce@us.es

ABSTRACT: We take as starting point the context logic studied in (Grossi, Meyer, and Dignum 2008) and (Aucher et al. 2009), which was concerned with the development of a dynamic logic of rules. Here we focus on issues of Formal Epistemology and omit the addition of deontic operators. We define some basic logics that are useful for the study of epistemic actions such as expanding, contracting or revising a theory taking into account different kinds of inferential results.

Key words: context logic, context models, modal operators, finite contexts.

1 Introducción

Dada una teoría, su extensión en un contexto se puede considerar como una ampliación de la misma con resultados obtenidos en procesos epistémicos, entre los que se cuentan los inferenciales, ya sean éstos deductivos, inductivos, abductivos o predictivos[1]. En general, las inferencias en un contexto de investigación científica son regidas por un sistema de razonamiento, el cual se puede identificar con lo que se ha dado en llamar lógica subyacente de la práctica correspondiente. Tal *lógica subyacente* es codificable en un sistema formal -un lenguaje formal, interpretable mediante cierta teoría semántica, y un mecanismo deductivo-.

No es habitual tomar en consideración una lógica subyacente mediante la cual se puedan captar procesos de inferencia no deductivas. En este trabajo proponemos unos sistemas de lógica básicos a partir de los cuales son definibles nuevas reglas inferenciales con vistas a representar operaciones epistémicas resultantes de procesos inferenciales no deductivos.

El modelo AGM de cambio epistémico, establecido a partir de (Alchourrón, Gärdenfors, and Makinson 1985), está basado en lógica clásica estándar y considera que el agente epistémico es omnisciente, es decir, que su base de conocimiento es cerrada bajo consecuencia lógica: para el lenguaje L y la teoría $\Theta \subset L$, si $Cn(\Theta) = \{\varphi \in L : \Theta \models \varphi\}$, entonces $\Theta = Cn(\Theta)$.

[1]Estudios pertinentes acerca de la abducción, y, en parte, la inducción, se hallan, entre otros, en (Aliseda 2006), (Gabbay and Woods 2006) y (Hintikka 1998); la noción de preducción se establece en (Rivadulla 2010).

Organizamos el trabajo de la siguiente manera. A esta introducción le sigue el apartado 2, en el que se presenta qué se entenderá por modelo de contexto y una lógica básica que es correcta y completa respecto de la clase de tales modelos. En el siguiente apartado se estudian mínimamente otros modelos -modelos de contextos no vacíos y el caso de modelos finitos-, y se define la lógica básica correspondiente. En un último apartado se incluyen algunas observaciones finales y se señalan nuevas líneas de trabajo, para finalizar con una lista de referencias.

2 Modelos contextuales

En una práctica científica se pueden presentar diversos estados epistémicos. Por ello, haciendo uso de una semántica kripkeana, ciertos conjuntos de estados o mundos posibles serán relevantes (o no) para cada contexto. Definimos entonces un *modelo de contextos*, o modelo de subconjuntos, de acuerdo con el siguiente clausulado:

1. P es un conjunto (no vacío) de variables proposicionales.
2. $\mathscr{C} = \{U\} \cup \{X, Y, Z, \dots\}$, es el conjunto de los contextos (realmente un conjunto de parámetros modales), donde U representa el operador *global*.

Entonces, se define un modelo de subconjuntos (contextos) $\mathscr{M} = (W, R, \mathscr{I})$, de manera que:

1. $W \neq \emptyset$ es el conjunto de mundos (o estados) posibles.
2. $R(U) = W$, y para cada $X \in \mathscr{C}$, si $X \neq U$, entonces $R(X) \subseteq W.R(X)$ representa el conjunto de mundos o estados *relevantes* para el contexto X. No obstante, por comodidad, y siempre que no se dé lugar a otras ambigüedades, llamaremos contexto al propio $R(X)$, es decir, se tomará al conjunto de mundos relevantes para el "contexto" como explicitación del contexto mismo. Obviamente, para el contexto global todos los estados son relevantes.
3. $\mathscr{I}$ es una valoración tal que para cada $p \in P, \mathscr{I}(p) \subset W.\mathscr{I}(p)$ representa el conjunto de estados que satisfacen p.

Se define un lenguaje proposicional a partir de P de acuerdo con la siguiente regla BNF:

$$\varphi := p | \neg\varphi | \varphi \to \psi | [X]\varphi$$

Naturalmente, $p \in P$. La expresión $[X]\varphi$ representa "en el contexto X vale φ". Dicho de otro modo, al anteponer el operador $[X]$ a la fórmula φ, se está expresando que "el contexto X es

necesariamente relevante para la fórmula φ". El operador $\langle X \rangle$ es el operador dual, de manera que $\langle X \rangle \varphi = \neg[X]\neg\varphi$ expresa que "en el contexto X es posible φ" (lo que es equivalente a afirmar que "no es el caso que en el contexto X no valga φ", o bien que "el contexto X es posiblemente relevante para la fórmula φ"). Conviene distinguir dos clases de fórmulas en relación con los operadores indicados, a saber:

1. *Contextuales*, en las cuales al menos un símbolo no lógico está bajo el alcance de un operador contextual, como, por ejemplo, en $[X]\varphi, [X]\varphi \to [Y]\psi, \langle X \rangle \varphi \to [Y] \langle X \rangle \varphi$, etc.

2. *Proposicionales*, en las cuales no ocurre ningún operador de contexto.

Dado un modelo contextual $\mathscr{M} = (W, R, \mathscr{I})$, para $w \in W$, se establece la noción de satisfacción, $\models$, según las siguientes cláusulas:

1. Para $p \in P, \mathscr{M}, w \models p$ si y sólo si (en adelante "syss") $w \in \mathscr{I}(p)$.

2. $\mathscr{M}, w \models \neg\varphi$ syss $\mathscr{M}, w \not\models \varphi$.

3. $\mathscr{M}, w \models \varphi \to \psi$ syss $\mathscr{M}, w \not\models \varphi$ o $\mathscr{M}, w \models \psi$.

4. Para $X \in \mathscr{C}, \mathscr{M}, w \models [X]\varphi$ syss para todo $w' \in R(X), \mathscr{M}, w' \models \varphi$. En este caso, en efecto, $\mathscr{M}$ en el estado w satisface "φ vale en el contexto X", pues en cada estado relevante para el contexto, el modelo satisface la fórmula.

Cuando $\mathscr{M}, w \models \varphi$ para todo $w \in W$, entonces φ es válida en $\mathscr{M}$, lo que se expresa como $\mathscr{M} \models \varphi$. Por otra parte, si para cada modelo contextual $\mathscr{M}$ se verifica que $\mathscr{M} \models \varphi$ (es decir, φ válida en $\mathscr{M}$), entonces φ es universalmente válida (en el sentido de los modelos contextuales), en símbolos $\models \varphi$.

El valor semántico de cada fórmula φ en el modelo $\mathscr{M}$ se establece siguiendo el clausulado de la regla BNF citada:

1. $\|p\|_{\mathscr{M}} = \mathscr{I}(p)$.

2. $\|\neg\varphi\|_{\mathscr{M}} = W \setminus \|\varphi\|_{\mathscr{M}}$.

3. $\|\varphi \to \psi\|_{\mathscr{M}} = \|\neg\varphi\|_{\mathscr{M}} \cup \|\psi\|_{\mathscr{M}}$.

4. $\|[X]\varphi\|_{\mathscr{M}} = R(X) \cap \|\varphi\|_{\mathscr{M}}$.

Nótese que, de acuerdo con las definiciones, si $\mathscr{M}, w \models [X]\varphi$, entonces el conjunto de (lo que podríamos llamar) los 'estados-φ' contiene el conjunto de estados relevantes para el contexto X, es decir, $R(X) \subseteq \|\varphi\|_{\mathscr{M}}$ y, en consecuencia, $R(X) \cap \|\varphi\|_{\mathscr{M}} = R(X)$.

Para contar con una lógica plena debemos definir un cálculo lógico, que por comodidad representamos mediante $\vdash$, de manera que $\vdash \varphi$ indica que la fórmula φ es demostrable en el cálculo, y para el conjunto de fórmulas Γ, $\Gamma \vdash \varphi$ expresa que φ se demuestra en el cálculo a partir de Γ. Este cálculo consta de los siguientes axiomas y reglas:

1. Todos los axiomas y reglas del cálculo proposicional clásico (ya sea tipo Hilbert, Lukasiewicz, Gentzen, etc.).
2. $[X]\varphi \to [Y][X]\varphi$.
3. $\langle X \rangle \varphi \to [Y] \langle X \rangle \varphi$.
4. $[U]\varphi \to \varphi$.
5. $[X](\varphi \to \psi) \to ([X]\varphi \to [X]\psi)$.
6. Regla de necesitación: de $\vdash \varphi$ se infiere $\vdash [X]\varphi$.

Sea $\mathfrak{M}$ la clase de los modelos contextuales. Respecto de esta clase, el cálculo es correcto y completo[2], es decir, para $\Theta \subset L$ y $\varphi \in L$, se verifica que:

$$\Theta \vdash \varphi \text{ syss } \Theta \models \varphi.$$

3 Clases específicas de contextos

3.1 Contextos no nulos

Surgen algunas cuestiones con vistas a las posibles aplicaciones de estas lógicas en ámbitos epistemológicos. Una de ellas es: ¿qué representa el contexto X en un modelo sin estados relevantes para el mismo? Sea $\mathscr{M} = (W, R, \mathscr{I})$ tal que para un $X \in \mathscr{C}$ se tiene que $R(X) = \emptyset$. En este caso tendremos que en el contexto X vale cualquier fórmula: $\mathscr{M}, w \models [X]\varphi$.

La consideración de que, en general, si una fórmula vale en un contexto, entonces dicha fórmula es posible en tal contexto, precisa que estemos hablando de contextos no vacíos. Un modelo tal que $R(X) = \emptyset$ no satisface $[X]\varphi \to \langle X \rangle \varphi$, al satisfacer al antecedente pero no al consecuente. Así pues, sea la clase de modelos $\mathfrak{M}_P = \{\mathscr{M} \in \mathfrak{M} : \mathscr{M} = (W, R, \mathscr{I}) \& \forall X \in \mathscr{C}, R(X) \neq \emptyset\}$. Respecto de

[2]En (Aucher et al. 2009) se presenta la prueba de corrección y completud del cálculo que definen, el cual es básicamente coincidente con el aquí presentado.

esta clase, al ser una subclase de la clase de los modelos de contexto, el cálculo es también correcto y completo.

3.2 Contextos finitos

Sea la clase de contextos finitos (no vacíos)

$$\mathfrak{M}_{PF} = \{\mathscr{M} \in \mathfrak{M} : \mathscr{M} = (W, R, \mathscr{I}) \& |W| \leq \omega_0 \ \& \ \forall X \in \mathscr{C}, R(X) \neq \emptyset\}$$

Entonces, dado un parámetro modal X, se puede hallar una fórmula ψ_X tal que su valor semántico es $R(X)$, es decir $\|\psi_X\|_{\mathscr{M}} = R(X)$. Así, por ejemplo, para el contexto global U, la fórmula podría ser $p \vee \neg p$, de tal manera que para el modelo $\mathscr{M} = (W, R, \mathscr{I}), \|p \vee \neg p\|_{\mathscr{M}} = R(U) = W$.

La correspondiente cláusula definitoria de la relación de satisfacción puede ahora adaptarse a esta circunstancia:

- Para $X \in \mathscr{C}, \mathscr{M}, w \models [X]\varphi$ syss para todo $w' \in R(X), \mathscr{M}, w' \models \varphi$ syss $\mathscr{M} \models \psi_X \to \varphi$.

Un nuevo cálculo es definible en relación a esta clase de modelos. Se trata de una simplificación del anterior y consta de los siguientes axiomas y reglas:

1. Todos los axiomas y reglas del cálculo proposicional clásico (ya sea tipo Hilbert, Lukasiewicz, Gentzen, etc.).
2. $[U]\varphi \to \varphi$.
3. Regla de necesitación: de $\vdash \varphi$ se infiere $\vdash [U]\varphi$.

Hemos de tener en cuenta que $[X]\varphi \to \langle X \rangle \varphi$ vale en cada modelo, dado que en éstos, por definición, se verifica que $R(X) \neq \emptyset$. Esta circunstancia es expresable en el siguiente esquema proposicional válido:

$$(\psi_X \to \varphi) \to (\psi_X \to \psi_X \wedge \varphi).$$

La regla de necesitación respecto de cualquier parámetro X es derivada. En efecto, se puede obtener la siguiente derivación:

1. $\vdash \varphi$ — Premisa.
2. $\vdash \varphi \to (\psi_X \to \varphi)$ — Axioma proposicional (para cualquier contexto X).
3. $\vdash \psi_X \to \varphi$ — *Modus ponens*, 1,2.

4. $\vdash [X]\varphi$ Definición-equivalencia 3.

Los axiomas 2, 3 y 5 del cálculo anterior, corresponden a los siguientes esquemas proposicionales, respectivamente:

2'. $(\psi_X \to \varphi) \to (\psi_Y \to (\psi_X \to \varphi))$.

3'. $\psi_X \wedge \varphi \to (\psi_Y \to \psi_X \wedge \varphi)$.

5'. $(\psi_X \to (\varphi \to \psi)) \to ((\psi_X \to \varphi) \to (\psi_X \to \psi))$.

Como hemos visto, la clase de los modelos finitos es una subclase de los modelos de contexto, es decir $\mathfrak{M}_{PF} \subseteq \mathfrak{M}$. Por otra parte, el cálculo es una adaptación del cálculo estudiado anteriormente. En consecuencia, el nuevo cálculo es correcto y completo respecto de $\mathfrak{M}_{PF}$, es decir:

$$\Theta \vdash \text{ syss } \Theta \models \varphi.$$

4 Observaciones finales

Para trabajar con lenguajes epistémicos otra manera de definir modelos contextuales tiene en cuenta la(s) correpondiente(s) relación(es) de accesibilidad[3]. En estos casos el lenguaje contiene un operador de conocimiento **K**, de manera que dada la fórmula φ, $\mathbf{K}\varphi$ expresa que φ es conocida.

Ahora un modelo contextual $M = (W, R, R_K, \mathscr{I})$, donde $W \neq \emptyset, R : \mathscr{C} \times W \longmapsto W$ -siendo $R(W, w)$ el conjunto de estados (mundos) relevante al mundo w en el contexto X-, R_K es la relación de accesibilidad e $\mathscr{I}$ la función interpretación. La noción de satisfacción se define a partir de las cláusulas siguientes:

1. $M, X, w \models p$ syss $\forall w' \in R(X, w), w' \in \mathscr{I}(p)$.

2. $M, X, w \models \mathbf{K}\varphi$ syss $\forall w' \in W$ tal que $(w, w') \in R(X, w)$, se verifica $M, X, w \models \varphi$.

Estas lógicas de contexto, especialmente cuando se toman modelos finitos sin contexto nulo, son aplicables en estudios de Epistemología Formal. El modelo AGM propone las operaciones de expansión, contracción y revisión, pero el sujeto epistémico es tomado como omnisciente, lo que lo aleja de una representación más acorde con las capacidades de los agentes reales.

Sin embargo, si se considera que una teoría se establece como resultado de una práctica científica en un contexto, la lógica subyacente debería tomarse como una lógica contextual ampliada para

[3]En (Rebuschi and Lihoreau 2009) se plantean modelos teniendo en cuenta un conjunto de agentes, por ello con relaciones de accesibilidad para cada agente; la semántica no coincide exactamente con la esbozada en nuestro trabajo.

comprender ciertas operaciones epistémicas. Refiriéndonos a las más básicas estudiadas, a partir de un lenguaje L y un modelo de contextos no nulos $\mathscr{M} = (W, R, \mathscr{I})$, se podría definir una operación de ampliación inferencial de una teoría Θ de la siguiente manera:

$$\Theta + \varphi = \{\psi \in L : \Theta \cup \varphi \hookrightarrow_X \psi\},$$

siempre que "$\hookrightarrow_X$", relación de 'consecuencia lógica en un contexto X', sea reflexiva y transitiva[4] y ψ sea el resultado de una deducción, o bien de una inducción, de una abducción o de una preducción.

Agradecimientos

Este trabajo se ha realizado en el marco de ejecución de los proyectos *Interpretaciones Alternativas de Lógicas no Clásicas* (referencia HUM5844 de la Consejería de Economía, Innovación y Ciencia de la Junta de Andalucía), y *Conciencia, Lógica y Computación* (referencia FFI2011-29609-C02-01, del Ministerio de Ciencia e Innovación).

REFERENCIAS

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*(2), 510–530.

Aliseda, A. (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Springer. Volume 330 of Synthese Library Series.

Aucher, G., Grossi, D., Herzig, A., & Lorini, E. (2009). *LORI 2009*, chapter Dynamic Context Logic, (pp. 15–26). Springer, LNAI 5834.

Gabbay, D. & Woods, J. (2006). *A Practical Logic of Cognitive Systems, vol. 2: The Reach of Abduction: Insight and Trial*. Elsevier.

Grossi, D., Meyer, J. J. C., & Dignum, F. (2008). The many faces of counts-as: A formal analysis of constitutive rules. *Journal of Applied Logic*, *6*, 192–217.

Hintikka, J. (1998). What is abduction? the fundamental problem of contemporary epistemology. *Transactions of the Charles S. Peirce Society*, *34*(3), 503–533.

Rebuschi, M. & Lihoreau, F. (2009). Contextual epistemic logic.

Rivadulla, A. (2010). Complementary strategies in scientific discovery: Abduction and preduction. In Pietarinen, A. V., Bergman, M., Paavola, S., & Rydenfelt, H. (Eds.), *Ideas in Action: Proceedings of the Applying Peirce Conference*, (pp. 264–276).

van Benthem, J. (2011). *Logical Dynamics of Information and Interaction*. Cambridge University Press.

[4]Son los requisitos mínimos necesarios para que una relación entre conjuntos de fórmulas y fórmulas sea considerada una relación de consecuencia lógica, ya sea en perspectiva sintáctica o semántica; es decir, para que estemos ante una lógica. Por otra parte, la Lógica debe ocuparse de estudiar los flujos de información (véase (van Benthem 2011)).

UNA LÓGICA EPISTÉMICA BIDIMENSIONAL

JOSÉ P. ÚBEDA RIVES
UNIVERSITAT DE VALÈNCIA

ABSTRACT: We analyse a propositional epistemic logic (LAK), which was described by A. J. Grove and J. Y. Halpern, and it is expressive enough to cope with anonymous agents, evolving groups of agents, and relative reference. o y palabras clave en inglés.

Key words: Bidimensional Modal Logic, Epistemic Logic, Knowledge, Lógica modal bidimensional, Multi-agent Systems

1 Introducción

Grove & Halpern (1993) estudian la lógica del conocimiento en el ámbito de la lógica proposicional y desean dar cuenta al menos de las siguientes situaciones

- El número de agentes no se conoce y puede variar de mundo a mundo.
- La composición de un grupo puede variar de mundo a mundo y su nombre no es rígido.
- Un agente no tiene que saber su nombre, ni si pertenece a un grupo.
- Pueden existir agentes anónimos.
- Existen nombres relativos y conocimiento sobre autoidentidad.

En esta comunicación vamos a analizar la lógica más general, **LAK**, que estos autores proponen para dar cuenta de esas situaciones. Para ello, describimos dicha lógica (sección 2) y comentamos sus propiedades (sección 3).

2 Descripción de la lógica

A. J. Grove y J. Y. Halpern proponen como la estructura más general, en el estilo de los mundos posibles de Kripke, la **AK**, que denominan estructura de *mundos posibles con conocimiento de la*

autoidentidad, sobre Φ, un conjunto de enunciados atómicos, y $\mathscr{N}$, un conjunto de nombres para grupos de agentes:

$$\mathfrak{M} = \langle \mathscr{W}, \mathscr{A}, \alpha, \mathscr{K}, \pi, \mu \rangle$$

donde

1. $\mathscr{W}$ es un conjunto de *mundos*, los mundos *posibles*;

2. $\mathscr{A}$ es un conjunto de *agentes*, los agentes *posibles*;

3. $\alpha : \mathscr{W} \to \wp(\mathscr{A})$, de forma que $\alpha(w)$ es el conjunto de agentes *vivos* en el mundo w; es el mecanismo que permite modelar la variación del número de agentes;

4. $\mathscr{K}$ es una relación sobre $\mathscr{W} \times \mathscr{A}$, tal que $\langle w, a \rangle \mathscr{K} \langle w', a' \rangle$ significa que para el agente a en el mundo w es *posible* el mundo w', donde *a se ve como* el agente a'; es uno de los mecanismos para modelar el conocimiento de un agente o grupo de agentes;

5. $\mu : \mathscr{W} \times \mathscr{N} \to \wp(\mathscr{A} \times \mathscr{A})$, tal que $a\mu(w, \mathbf{n})b$ si el agente a esta en la relación $\mu(w, \mathbf{n})$ con el agente b;[1] es decir, para a en el mundo w, *b tiene el nombre* $\mathbf{n}$. $\{b | \langle a, b \rangle \in \mu(w, \mathbf{n})\}$ es el grupo de agentes que a llama $\mathbf{n}$ en el mundo w; es el mecanismo para modelar la composición de los grupos de agentes;

6. $\pi : \mathscr{W} \to (\mathscr{A} \to (\Phi \to \{\mathbf{V}, \mathbf{F}\}))$ es tal que $\pi(w)(a)(p)$ es verdadero $\mathbf{V}$ o falso $\mathbf{F}$, es decir, el enunciado atómico *p en el mundo w bajo la perspectiva del agente a* es $\mathbf{V}$ o $\mathbf{F}$.

El lenguaje que proponen para las lógicas basadas en estas estructuras se construye sobre los siguientes símbolos:

1. las conectivas de la lógica de enunciados;

2. los símbolos en Φ para las fórmulas atómicas;

3. para cada $\mathbf{n} \in \mathscr{N}$, los símbolos $S_{\mathbf{n}}$ y $E_{\mathbf{n}}$, que representan "alguien en el grupo $\mathbf{n}$ conoce" y "todos en el grupo $\mathbf{n}$ conocen";[2]

4. el símbolo K_I, que representa "yo conozco";[3]

5. signos de puntuación.

[1] Usaremos $\langle a, b \rangle \in \mu(w, \mathbf{n})$ como una forma equivalente de $a\mu(w, \mathbf{n})b$.

[2] Aunque no los estudian, Grove y Halpern mencionan los operadores $C_{\mathbf{n}}$ y $D_{\mathbf{n}}$ que representan "el *conocimiento común*" y el "conocimiento distribuido" en el grupo $\mathbf{n}$.

[3] En (Grove & Halpern, 1993, p. 362) se introduce el nombre I de forma que $a\mu(w, I)a'$ si y sólo si $a = a'$; es decir, I se refiere siempre a un único agente. Es un mecanismo para modelar la referencia relativa.

La semántica se basa sobre la relación $\models$, tal que $(w,a) \models \varphi$ se tiene, si *desde la perspectiva del agente a en el mundo w*, φ es *verdadera*. Su definición es recursiva:

1. si $\varphi \in \Phi$, $(w,a) \models \varphi$ si y sólo si $\pi(w)(a)\varphi = \mathbf{V}$;

2. si $\varphi = \neg\psi$, $(w,a) \models \varphi$ si y sólo si no se tiene $(w,a) \models \psi$;

3. si $\varphi = \chi \wedge \psi$, $(w,a) \models \varphi$ si y sólo si $(w,a) \models \chi$ y $(w,a) \models \psi$;

4. si $\varphi = E_{\mathbf{n}}\psi$, $(w,a) \models \varphi$ si y sólo si

$$\forall b \in \mathscr{A}(\langle a,b\rangle \in \mu(w,\mathbf{n}) \rightarrow (\forall w' \in \mathscr{W})(\forall b' \in \mathscr{A})(\langle w,b\rangle \mathscr{K} \langle w',b'\rangle \rightarrow (w',b') \models \psi));$$

5. si $\varphi = S_{\mathbf{n}}\psi$, $(w,a) \models \varphi$ si y sólo si

$$\exists b \in \mathscr{A}(\langle a,b\rangle \in \mu(w,\mathbf{n}) \wedge (\forall w' \in \mathscr{W})(\forall b' \in \mathscr{A})(\langle w,b\rangle \mathscr{K} \langle w',b'\rangle \rightarrow (w',b') \models \psi));$$

6. si $\varphi = K_I$, $(w,a) \models \varphi$ si y sólo si[4]

$$(\forall w' \in \mathscr{W})(\forall b \in \mathscr{A})(\langle w,a\rangle \mathscr{K} \langle w',b\rangle \rightarrow (w',b) \models \psi).$$

3 Comentarios sobre LAK

Sobre esta lógica podemos hacer los siguientes comentarios:

- Las estructuras **AK** son bidimensionales heterogéneas, ya que se definen a partir de dos universos.[5] Por ello, permite introducir dos operadores de necesidad $\boxminus$ y $\boxminus$, tales que $(w,a) \models \boxminus\varphi$ si y sólo si φ es verdadero en w sin depender de la perspectiva de los agentes:[6]

$$(w,a) \models \boxminus\varphi \Leftrightarrow \forall b \in \mathscr{A}((w,b) \models \varphi)$$

  y $(w,a) \models \boxminus\varphi$ si y sólo si φ es verdadero en todos los mundos desde la perspectiva de a:

$$(w,a) \models \boxminus\varphi \Leftrightarrow \forall y \in \mathscr{W}((y,a) \models \varphi).$$[7]

[4]Esta equivalencia se sigue del hecho de que I es un nombre que siempre denota un único agente y de que $\mathbf{K}_I$ es lo mismo que $\mathbf{E}_I$ y que $\mathbf{S}_I$.

[5]Para un estudio de las lógicas modales multidimensionales puede consultarse (Marx & Venema, 1997).

[6]Los enunciados φ, tales que $(w,a) \models \boxminus\varphi$, son los enunciados *absolutos*, su valor de verdad depende sólo del mundo posible.

[7]La lógica de estos operadores es la lógica **S5** para cada uno de ellos junto con el axioma $\boxminus\boxminus\varphi \leftrightarrow \boxminus\boxminus\varphi$.

- La función α da lugar a la posibilidad de introducir un operador modal constante **v** tal que

$$(w,a) \models \mathbf{v} \Leftrightarrow a \in \alpha(w);$$

es decir, **v** se tiene en el mundo w bajo la perspectiva de a si y sólo si a vive en w. Si introducimos como axioma el enunciado **v** podemos caracterizar las estructuras en las que $\alpha(w) = \mathscr{A}$ para todo $w \in \mathscr{W}$; es decir, aquellas en las que el número de agentes que existe en cada mundo es siempre el mismo.

- La relación $\mathscr{K}$ da lugar a un operador modal normal **K**, tal que

$$(w,a) \models \mathbf{K}\varphi \Leftrightarrow \forall y \in \mathscr{W} \forall b \in \mathscr{A} (\langle w,a\rangle \mathscr{K} \langle y,b\rangle \rightarrow (y,b) \models \varphi).$$

K sería un operador de conocimiento de un "agente" cuando el conjunto de los mundos posibles es $\mathscr{W} \times \mathscr{A}$ y $\mathscr{K}$ la relación de *accesibilidad* y coincide con el operador $\mathbf{K}_I$ (Grove & Halpern, 1993, p. 364).

- ¿Qué propiedades debe tener $\mathscr{K}$ para que **K** o $\mathbf{K}_I$ sea un operador de conocimiento? En (Grove & Halpern, 1993, p. 361) se propone que $\mathscr{K}$ sea tal que

$$\forall w, w' \in \mathscr{W} \forall a, a' \in \mathscr{A} (\langle w,a\rangle \mathscr{K} \left\langle w', a'\right\rangle \rightarrow a \in \alpha(w) \wedge a' \in \alpha(w')).$$

En este caso, se cumple el axioma

$$\neg\mathbf{v} \rightarrow \mathbf{K}\bot;$$

es decir, si un agente a no vive en un mundo w, entonces $(w,a) \models \mathbf{K}_I\varphi$, para cada fórmula φ, a conoce "todo".[8] La única forma de conseguir que $\mathbf{v} \rightarrow \mathbf{K}\bot$ sea siempre falso es exigir que se cumpla

$$\forall w \in \mathscr{W} \forall a \in \mathscr{A} (a \in \alpha(w) \rightarrow \exists w' \in \mathscr{W} \exists a' \in \mathscr{A} (\langle a,w\rangle \mathscr{K} \left\langle a', w'\right\rangle)) \qquad (A).$$

Los autores proponen que $\mathscr{K}$ sea una relación de equivalencia sobre el conjunto $\{\langle w,a\rangle \,|a \in \alpha(w)\}$, lo que implica que cumple (A).[9]

- En (Grove & Halpern, 1993, p. 364) se dice que es válido el enunciado $\mathbf{K}_I\varphi \rightarrow \varphi$; eso quiere decir que $\mathscr{K}$ es reflexiva sobre $\mathscr{W} \times \mathscr{A}$, lo que implica que $\alpha(w) = \mathscr{A}$, para cada $w \in \mathscr{W}$.

[8] Lo que conoce un agente a en el mundo w podemos identificarlo con el conjunto $\{\varphi | (w,a) \models \mathbf{K}_I\varphi\}$.
[9] Exigiendo estas propiedades a $\mathscr{K}$ se tiene que $\mathbf{v} \leftrightarrow \neg\mathbf{K}\bot$. En este caso se puede eliminar el operador **v**.

¿Para qué introducen la función α en su estructura, si no sirve para nada? Lo lógico habría sido establecer en lugar de dicho axioma, el axioma

$$\mathbf{v} \wedge \mathbf{K}_I \varphi \rightarrow \varphi.$$

- La relación μ da lugar, para cada $\mathbf{n} \in \mathscr{N}$, a un operador modal constante $\mathbf{c_n}$, tal que

$$(w,a) \models \mathbf{c_n} \Leftrightarrow a\mu(w,\mathbf{n})a;$$

es decir se tiene $\mathbf{c_n}$ en el mundo w bajo la perspectiva de a, si a es un $\mathbf{n}$.[10]

- Si para un $w \in \mathscr{W}$, $a \in \mathscr{A}$ y $\mathbf{n} \in \mathscr{N}$ no existe ningún $b \in \mathscr{A}$ tal que $a\mu(w,\mathbf{n})b$, entonces se tiene, para todo enunciado φ, $(w,a) \models \mathbf{E_n}\varphi$ y $(w,a) \models \neg\mathbf{S_n}\varphi$. Estos hechos sugieren la introducción, para cada $\mathbf{n} \in \mathscr{N}$, de un operador modal constante $\mathbf{n}_\emptyset$ con la semántica siguiente:

$$(w,a) \models \mathbf{n}_\emptyset \Leftrightarrow \neg\exists x \in \mathscr{A}(a\mu(w,\mathbf{n})x).$$

Así, sería válido, para cada enunciado φ, el siguiente axioma: $\mathbf{n}_\emptyset \rightarrow \mathbf{E_n}\varphi \wedge \neg\mathbf{S_n}\varphi$.

- En (Grove & Halpern, 1993, p. 361) se exige respecto a la relación μ que, para todo $w \in \mathscr{W}$ y $\mathbf{n} \in \mathscr{N}$, se cumpla

$$\forall a,b \in \mathscr{A}(a\mu(w,\mathbf{n})b \Rightarrow a,b \in \alpha(w)).$$

Esta exigencia elimina la posibilidad de trabajar con nombres que puedan aplicarse a agentes que no viven en un mundo, pero hace que sea válido el enunciado $\neg\mathbf{S_n}\bot$.[11]

- Si no aceptamos la exigencia de Grove y Halpern relativa a la relación μ y aceptamos que sea posible que se tenga $a\mu(w,\mathbf{n})b$ aunque los agentes a y b no vivan en w, podemos introducir, para cada $\mathbf{n} \in \mathscr{N}$, un operador modal constante $\mathbf{n}_\theta$ con la siguiente semántica:

$$(w,a) \models \mathbf{n}_\theta \Leftrightarrow \exists x \in \mathscr{A}(a\mu(w,\mathbf{n})x \wedge x \notin \alpha(w)).$$

Así, sería válido, para cada enunciado φ, el siguiente enunciado $\mathbf{n}_\theta \rightarrow \mathbf{S_n}\varphi$, siempre que, para cada enunciado φ, se cumpliera $\neg\mathbf{v} \rightarrow \mathbf{K}_I\varphi$.

- Para poder establecer que un agente que no vive en un mundo w no conoce nada, es decir, su

[10] En (Grove & Halpern, 1993, p. 366) se introduce $\mathbf{c_n}$ utilizando el nombre $\mathbf{n}$ como nombre y como enunciado.
[11] Los autores lo introducen como el axioma $S1'$ (Grove & Halpern, 1993, p. 364).

conocimiento es vacío debe modificarse la semántica de $\mathbf{K}_I$, $\mathbf{E_n}$ y $\mathbf{S_n}$ en la siguiente forma:

$$(w,a) \models \mathbf{K}_I\psi \Leftrightarrow \mathbf{v} \wedge (\forall w' \in \mathscr{W})(\forall b \in \mathscr{A})(\langle w,a\rangle \mathscr{K} \langle w',b\rangle \rightarrow (w',b) \models \psi);$$

$$(w,a) \models \mathbf{E_n}\psi \Leftrightarrow \forall b \in \mathscr{A}(a\mu(w,\mathbf{n})b \rightarrow (w,b) \models \mathbf{K}_I\psi);$$

$$(w,a) \models \mathbf{S_n}\psi \Leftrightarrow \exists b \in \mathscr{A}(a\mu(w,\mathbf{n})b \rightarrow (w,b) \models \mathbf{K}_I\psi).$$

Con esta nueva semántica, para cada enunciado φ, serían válidos $\neg\mathbf{v} \rightarrow \neg\mathbf{K}_I\varphi$ y $\mathbf{n}_\theta \rightarrow \neg\mathbf{E_n}\varphi$.[12]

¿Puede ser vacío alguno de los conjuntos $\mathscr{W}$, $\mathscr{A}$ o $\mathscr{N}$? En los dos primeros casos la relación $\models$ definida sería vacía, ya que no habría ningún par $\langle w,a\rangle$. En el tercer caso, no tendríamos símbolos de la forma $S_\mathbf{n}$ ni de la forma $E_\mathbf{n}$ y la lógica obtenida sería la de proposiciones.

Pero, es usual definir a partir de $\models$ nuevas relaciones que usando los operadores $\boxminus$ y $\boxplus$ y sus duales $\Diamond\!\equiv \neg\boxplus\neg$ y $\Diamond\!\equiv \neg\boxminus\neg$ son las siguientes

1. $w \models_\forall \varphi \leftrightarrow \forall a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \boxplus\varphi$;
2. $w \models_\exists \varphi \leftrightarrow \exists a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \Diamond\!\varphi$;
3. $a \models_\forall \varphi \leftrightarrow \forall w \in \mathscr{W}((w,a) \models \varphi) \leftrightarrow (w,a) \models \boxminus\varphi$;
4. $a \models_\exists \varphi \leftrightarrow \exists w \in \mathscr{W}((w,a) \models \varphi) \leftrightarrow (w,a) \models \Diamond\!\varphi$;
5. $\models_{\forall\forall} \varphi \leftrightarrow \forall w \in \mathscr{W}\forall a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \boxminus\boxplus\varphi$ (φ es *válida* en $\mathfrak{M}$);
6. $\models_{\exists\forall} \varphi \leftrightarrow \exists w \in \mathscr{W}\forall a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \Diamond\!\boxplus\varphi$;
7. $\models_{\forall\exists} \varphi \leftrightarrow \forall w \in \mathscr{W}\exists a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \boxminus\Diamond\!\varphi$;
8. $\models_{\exists\exists} \varphi \leftrightarrow \exists w \in \mathscr{W}\exists a \in \mathscr{A}((w,a) \models \varphi) \leftrightarrow (w,a) \models \Diamond\!\varphi$ (φ es *satisfacible* en $\mathfrak{M}$).

Ahora, si $\mathscr{W} = \mathscr{A} = \emptyset$, entonces, para toda fórmula φ, se tiene $\models_{\forall\exists} \varphi$, $\models_{\forall\forall} \varphi$, $\neg(w \models_{\exists\forall} \varphi)$ y $\neg(w \models_{\exists\exists} \varphi)$. Si $\mathscr{A} = \emptyset$ y $\mathscr{W} \neq \emptyset$, para toda φ, se tiene $w \models_\forall \varphi$ y $\neg(w \models_\exists \varphi)$. Si $\mathscr{W} = \emptyset$ y $\mathscr{A} \neq \emptyset$, para toda φ, se tiene $a \models_\forall \varphi$ y $\neg(a \models_\exists \varphi)$. Si realmente lo que interesa es $\models_{\forall\forall}$, no hay problema en que $\mathscr{W}$ o $\mathscr{A}$ sean vacíos. Y todas las estructuras **AK**, tales que $\mathscr{W}$ o $\mathscr{A}$ sean vacíos, se caracterizan por los axiomas $\vdash \mathbf{E_n}\varphi$, donde $\mathbf{n} \in \mathscr{N}$.

REFERENCIAS

[12]Un análisis de esta nueva semántica y de otras lógicas diferentes puede verse en (Ubeda Rives, 2004, 2005, 2008).

Grove, A. J. & Halpern, J. Y. (1993). Naming and identity in epistemic logics, part i: The propositional case. *Journal of Logic and Computation*, *3*(4), 345–378.

Marx, M. & Venema, Y. (1997). *Multi-Dimensional Modal Logic*. Kluwer Academic Publishers.

Ubeda Rives, J. P. (2004). Epistemic logics allowing different sets of agents in each world. In *Third International Workshop on the History and Philosophy of Logic, Mathematics, and Computation, HPLMC-04*, (pp. 169–174). Servicio Editorial Universidad del Pais Vasco, San Sebastian.

Ubeda Rives, J. P. (2005). Marcos de kripke y sistemas multiagentes. In Castillo Valdipieso, P. A. & Córdoba, A. (Eds.), *Actas del I Simposio de Sistemas Complejos*, (pp. 39–43). Paraninfo.

Ubeda Rives, J. P. (2008). Lógicas epistémicas y sistemas multiagentes. *Quaders de filosofia i ciència*, *38*, 7–21.

TIEMPO Y PUNTOS DE VISTA

MARGARITA VÁZQUEZ[1]
Universidad de La Laguna
mvazquez@ull.es

ABSTRACT: This paper analyzes the temporal structures that are appropriate to study the notion of point of view. It is examined how to understand time (McTaggart series) and how to formalize it (Prior temporal logic). It is also shown a logical approach to the notion of point of view (Hautamäki). The conclusion is that, in order to give a proper approach, aspects of these three contributions should be combined.
KEYWORDS: Point of view, temporal logic, time, Prior, McTaggart.

Introducción

En este trabajo se quiere hacer un primer intento de respuesta a la siguiente pregunta: ¿Cuáles son las estructuras temporales apropiadas para analizar conceptualmente la noción de punto de vista?

La noción de punto de vista está presente en muy diversas áreas de la filosofía actual. En todas estas áreas filosóficas la noción de puntos de vista siempre está estrechamente vinculada a la existencia de problemas a la hora de establecer relaciones epistémicas, semánticas, normativas y ontológicas con una realidad no dependiente de los sujetos. La necesidad de apelar a los puntos de vista al tratar cualquier fenómeno casi siempre se ha tomado como implicando la devaluación ontológica del planteamiento ofrecido, como si tal apelación sólo pudiera ser una admisión a efectos pragmáticos. Quedarían fuera de este diagnóstico dos áreas: 1) la filosofía del lenguaje más reciente en donde se ha ido asumiendo la dependencia del contexto, y 2) las teorías sobre la racionalidad, donde poco a poco se ha ido aceptando que la noción de racionalidad acotada es algo más que un sustituto provisional de otro tipo de racionalidad más deseable. Por supuesto, otras áreas se han sumado a este movimiento de reivindicación de las perspectivas.

[1] Este trabajo ha sido financiado con los proyectos FFI2008-01205. Puntos de Vista. Una Investigación Filosófica y FFI2011-24549. Puntos de vista y estructuras temporales.

En Liz, M. (ed.) (2012), Liz y Vázquez (2010), Liz y Vázquez (2011) y Vázquez y Liz (2011), hemos analizado la estructura, dinámica y modo peculiar de existencia de los puntos de vista. Los puntos de vista objeto de estudio son puntos de vista personales, con posibles contenidos tanto no conceptuales como conceptuales y con capacidades reflexivas.

Ahora se trataría de empezar a indagar los aspectos temporales de los puntos de vista, es decir explorar las estructuras temporales más relevantes a la hora de analizar la dinámica de los puntos de vista con cierto grado de complejidad. Para ello vamos a examinar cómo se puede entender el tiempo y cómo se puede formalizar lógicamente.

1. Concepciones del tiempo: McTaggart

McTaggart[2] es la referencia clásica respecto a dos maneras muy distintas de entender el tiempo[3]: la serie A y la serie B.

La serie A consistiría en situar los acontecimientos en un pasado, presente o futuro. La serie B los situaría de acuerdo a una relación de anterioridad-posterioridad. McTaggart defendió que las series de tipo A son esenciales para que exista propiamente el tiempo. Pero también argumentó McTaggart que, por ello mismo, el tiempo es irreal. Las series A son, según McTaggart, profundamente contradictorias, o cuando menos paradójicas.

Los planteamientos de McTaggart han configurado las discusiones posteriores, y siguen presentes en la actualidad. Así, suele distinguirse entre concepciones o teorías del tiempo de tipo A y de tipo B. Y sigue viva la idea de que una serie temporal de tipo A, aunque resulte esencial para poder tener propiamente un tiempo que "fluye" (un tiempo no simplemente espacializado), nos enfrenta a series problemas conceptuales, si la entendemos como algo más que un mero tiempo interno (subjetivo, epifenoménico, etc., y en último término irreal).

Siguiendo la terminología de McTaggart, en un primer acercamiento podríamos pensar que los puntos de vista complejos se podrían analizar como una diversidad de estructuras temporales de la serie A, ramificadas, combinadas y multimodalmente enriquecidas. De hecho, los planteamientos clásicos de lógica temporal, como el de Prior, se entienden habitualmente como ligados a estructuras de tipo serie A.

Estas estructuras temporales, ligadas a la serie A, definirían una clase de "tiempo interno" no inerte y homogéneo, sino fuertemente direccionado, estructurados y constitutivo respecto a los puntos de vista involucrados.

---

2 McTaggart (1927)
3 Sobre este tema véase Álvarez (1994).

2. La perspectiva de la lógica temporal: Prior

Prior es la referencia clave a la hora de hablar de la lógica temporal. Aunque se pueden rastrear antecedentes en la lógica estoica, Aristóteles y la lógica medieval, es en la obra de Prior cuando se empiezan a desarrollar sistemas de lógica temporal propiamente dichos. Su obra influye en todos los desarrollos de la misma a partir de los años 60, pues allí podemos encontrar el germen de prácticamente todos los sistemas que se siguen estudiando actualmente. Sus trabajos más relevantes son Prior (1957), Prior (1967) y Prior (1969)[4].

Una idea fundamental de Prior (inspirada en sus estudios sobre Diodoro de Cronos) es que nuestro lenguaje y nuestro pensamiento tienen una "perspectiva interna de tiempo" en la cual se representa el tiempo fijando un pasado y un futuro respecto a un ahora cambiante. Una lógica del tiempo debería respetar este sentido interno de una temporalidad que requiere un "ahora que fluye". Unos de los desafíos más importantes para el enfoque de Prior es deshacer la sensación de paradoja (denunciada justamente por McTaggart) que parece entrañar esta idea de que el propio tiempo "fluya".

Para construir sus lógicas, Prior introduce, entre otros, operadores tipo modal para referirse al pasado y al futuro. Así tenemos P ("alguna vez en el pasado"), H ("siempre en el pasado"), F ("alguna vez en el futuro") y G ("siempre en el futuro"). Estos operadores permiten construir proposiciones en las que se haga referencia al pasado, presente y futuro. Por ejemplo, si se quiere expresar que el tiempo fluye, se introduciría un axioma como $A \rightarrow HFA$, es decir, si se da A entonces ha sido siempre en el pasado que A sería alguna vez en el futuro. De esta forma, los acontecimientos quedan ordenados como pasados y futuros con respecto a un presente. La similitud de estos operadores con la serie A de McTaggart es manifiesta.

En general, estos sistemas definen un flujo de tiempo, un modelo temporal ***T*** como una estructura $(T,<,v)$, donde $T \neq \emptyset$ es un conjunto de momentos, $T=\{\ldots, t_{-2}, t_{-1}, t_0, t_1, t_2, t_3, \ldots\}$[5]; $< \subseteq T^2$ es la relación antes/después sobre los momentos y tiene diferentes propiedades dependiendo del flujo de tiempo, siendo siempre irreflexiva[6]; v es una asignación, siendo F el conjunto de fbf, $v: FXT \rightarrow \{1,0\}$. De aquí, podemos los subconjuntos de $V(p) \subseteq T$ en los que una fórmula atómica es

[4] Dadas las dificultades que entraña la obra de Prior, por la notación polaca que utiliza, se ha hecho una edición de Prior (1968) en Prior (2003) actualizando la notación.

[5] t_0 se toma normalmente como el momento presente o algún momento especial del tiempo.

[6] $(\forall t_i \in T \neg(t_i R t_i))$

verdadera. La asignación es definida por inducción, siendo la evaluación para las fórmulas propiamente temporales, para cualquier $t_i \in T$, $A,B \in F$ y variable proposicional p:

$v(FA, t_i)=1$ si y solo si $\exists t_j \in M\ (t_i < t_j \wedge v(A, t_j)=1)$

$v(PA, t_i)=1$ si y solo si $\exists t_j \in M\ (t_i > t_j \wedge v(A, t_j)=1)$

El propio Prior plantea, y se han desarrollado posteriormente, sistemas temporales bidimensionales, donde se combinan operadores temporales y modales, para poder dar cuenta del indeterminismo y de las ramificaciones del tiempo (en concreto, de las ramificaciones hacia el futuro). En estas lógicas la evaluación no será solo sobre los momentos de tiempo (t), como en la expuesta anteriormente, sino sobre dos índices, el momento y la historia o rama. En cualquier caso, para la evaluación se utiliza siempre una relación de ulterioridad, o relación antes/después, como la de la serie B de McTaggart.

Siguiendo esta misma notación y la combinación de operadores modales, Hautamäki planteó en 1984 una lógica proposicional de los puntos de vista.[7]

3. Otras aportaciones: Hautamäki

La lógica de los puntos de vista del filósofo finlandés Antti Hautamäki[8] consta de dos tipos de operadores, los clásicos de la lógica modal (*L* y *M*) y unos propios de esta nueva lógica, *A* y *R*. *A* se interpreta como "desde todos los puntos de vista" y *R* como "desde algún punto de vista". La semántica es bidimensional, como las del tiempo indeterminista, siendo la evaluación a dos índices, el mundo y el punto de vista. Hautamäki presenta distintas axiomatizaciones de esta lógica.

Estos sistemas definen un modelo de punto de vista ***M*** como una estructura (W,I,R,S,v), donde $W \neq \emptyset$ es un conjunto de mundos; $I \neq \emptyset$ es un conjunto de puntos de vista; R y S son relaciones en WXI, $R \subseteq (WXI)^2$ y $S \subseteq (WXI)^2$; v es una asignación, siendo F el conjunto de fbf, v: $FXWXI \rightarrow \{1,0\}$. La asignación es definida por inducción, siendo la evaluación para las fórmulas con operadores modales y de puntos de vista, para cualquier $w_i \in W$ y $i_i \in I$, $B \in F$ y variable proposicional p:

$v(LB, w_{i,}\ i_{i,})=1$ si y solo si $\forall w_j \in W\ (<w_{i,},i_i>R<w_j,i_i> \rightarrow v(B, w_j, i_i)=1)$

$v(AB, w_{i,}\ i_i)=1$ si y solo si $\forall i_j \in I\ ((<w_i,i_i>R<w_i,i_j> \rightarrow v(B, w_{i,}, i_j)=1)$

---

7 Hautamäki (1983).

8 Para Hautamäki, un punto de vista es una manera de conceptualizar el mundo, y sería equivalente a un esquema conceptual, un esquema lingüístico, un marco conceptual o una perspectiva teórica.

Hautamäki expone también la aplicación de esta lógica para la lógica temporal, definiendo *A* y *R* de forma similar a como Prior definía G y H. Esto tiene similitudes con algunos sistemas de tiempo ramificado. Lo que sucede es que los sistemas de Hautamäki no ponen ningún tipo de restricciones a las combinaciones entre W e I, con lo que las alternativas proliferarían por todas partes y no habría ninguna diferencia, en cuanto a la verdad, entre los hechos pasados y los futuros.

4. Conclusión

Se ha señalado muchas veces la afinidad de la lógica temporal tipo Prior con las concepciones del tiempo basadas en la serie A de McTaggart. Mi conclusión es que no se decanta por ninguna de las dos series, sino que conjuga ambas, dado que: (1) a nivel sintáctico se usan los operadores de pasado y futuro (serie A), y (2) en la semántica se utilizan relaciones de ulterioridad o relaciones antes/después (serie B).

De esta manera, cualquier enunciado se escribe mediante fórmulas que involucran las ideas de pasado, presente y futuro, pero a la hora de ser evaluadas dichas fórmulas se toma en consideración la relación de antes/después. Así, cuando evaluamos una fórmula, el ahora va cambiando y con él aquello que es considerado pasado o futuro (independientemente de lo que fuera el ahora al comenzar la evaluación). Lo que nos importa a la hora de hacer la evaluación es qué momentos van antes (o después) de otros. Y con estas evaluaciones, atribuimos la verdad o falsedad a unas fórmulas que, a nivel sintáctico, hacen referencia al pasado, al presente y al futuro.

Por lo tanto, a la hora de explorar la dinámica de los puntos de vista, las estructuras temporales que subyacen, no parece que tendríamos porqué elegir entre las series A y B de McTaggart, sino que ambas podrían coexistir. Y una lógica temporal tipo Prior parece un buen punto de partida.

Hemos visto que Hatumäki plantea unas lógicas para los puntos de vista, y, en base a ellos, define también unas lógicas temporales. Estas lógicas temporales tienen unas graves limitaciones. Por ello, quizás fuese más adecuado seguir el camino inverso, para la elaboración de una lógica de los puntos de vista, tomar como base una lógica del tiempo indeterminista y, sobre ella, definir unos operadores similares a los de Hautamäki.

REFERENCES

Álvarez, S. 1994, 'Tiempo, cambios e indeterminismo', *Análisis Filosófico*, 14(2):113-130.
Hautamäki, A. 1983, 'The Logic of Viewpoints', *Studia Logic: An International Journal for Symbolic Logic*, 42(2/3): 187-196.

Herrera, R. y Vázquez, M. 2011, 'Hacia una lógica temporal epistémica basada en lenguajes híbridos', *Análisis Filosófico*.
Liz, M. (ed.) 2009, *La realidad sin velos*, Barcelona, Laertes.
Liz, M. (ed) 2012, *Puntos de vista. Una investigación filosófica*, Barcelona, Laertes (en prensa).
Liz, M. y Vázquez, M. 2010, 'Subjects and points of view. Where priviledged access meets with intersubjectivity', en *DISCOS International Conference in Intersubjectivity and the Self*, Budapest.
Liz, M. y Vázquez, M. 2011, **'**Two approaches to the notion of point of view', *14th Congress of Logic, Methodology and Philosophy of Science,* Nancy, Francia
McTaggart, J.M.E. 1927, *The Nature of Existence, vol.II*, compilado por C.D. Broad, Massachusetts, Cambridge University Press.
Ohrstrom, P. y Hasle, P. 1995, *Temporal Logic. From Ancient Ideas to Artificial Intelligence*, Kluwer Academic Press.
Ponte, M. y Vázquez, M. 2008, 'An approach to the notion of point of view through hybrid temporal logic', *ESPP 2008*, Utrech.
Ponte, M. y Vázquez, M. 2012, 'Tense and Temporal Reference. Hybrid Temporal Logic', *Logique et Analyse*.
Prior, A. 1957, *Time and Modality*, Oxford, Clarendon Press.
Prior, A. 1967, *Past, Present and Future*, Oxford, Clarendon Press.
Prior, A. 1968, *Papers on Time and Tense*, Oxford, Clarendon Press.
Prior, A. 2003, *Papers in Time and Tense. New Edition***, e**ditado por Hasle, Ohrstrom, Braüner y Copeland, Oxford Universited Press.
Vázquez, M. y Liz, M. 2011, 'Models as Points of View. The Case of System Dynamics', *Foundations of Science* 16(4):383-391, Springer Verlag.

PHILOSOPHY OF LANGUAGE, PHILOSOPHY OF MIND, EPISTEMOLOGY

FILOSOFÍA DEL LENGUAJE, FILOSOFÍA DE LA MENTE, EPISTEMOLOGÍA

WITTGENSTEIN'S ANTI-DESCRIPTIVISM

JUAN JOSÉ ACERO, NEFTALÍ VILLANUEVA
Universidad de Granada

ABSTRACT: The aim of this paper is to show that Wittgenstein's anti-descriptivist approach to the meaning of mental states ascriptions is not restricted to a particular kind of first person mental states attributions –avowals. Our strategy is threefold. We will first provide textual evidence to make it apparent that Wittgenstein's non-relational treatment of mental states ascriptions is not restricted to avowals. Secondly, we will analyze three different arguments provided by Wittgenstein against the relational nature of mental states ascriptions. Finally, we will develop some ideas to try and make sense of the claim that by using third-person mental states ascriptions we are not describing, but expressing mental states.

KEYWORDS: Wittgenstein, descriptivism, avowals, mental-state attributions, expressivism

1. Varieties of anti-descriptivism

At least three uses of the term 'descriptivism' can be clearly distinguished in the literature, corresponding to different senses in which a certain theory can be said to be 'descriptivist':

i) Frege-Russell Theory: The meanings of proper names *are* sets of definite descriptions.

ii) Intensionalism: The meanings of proper names are *determined* by definite descriptions.

iii) Relationalism: mental states attributions are used to state a relation between a thinker and something else –a wish, a belief, some piece of knowledge, etc. (cfr. criticisms in Urmson, 1956; Kiteley 1964; Prior 1971; Recanati 2000, 2007, etc.)

Even though Wittgenstein opposes descriptivism in senses i) and ii) (PI 79), it is the third sense that is of interest for the purpose of this paper. Mental states attributions are not used to claim that any kind of relation is held between a thinker and an object, they are not used to describe a state of affairs, to represent 'a distribution in a space and time' (cfr. PI, ix).

In recent literature on the meaning of belief reports, arguments against relationalist views have been summarized in (Boër 2007, 39-65) and (Matthews 2007, 97-117). Matthews writes:

> 'even a cursory examination of the problems that plague the relational conception of belief should disabuse anyone of the notion that this assumption has any presumptive status. These problems have been carefully catalogued elsewhere (e.g., Richard 1990; Schiffer 1987, 1992), so I shall limit myself to a few brief reminders, which are intended to set the stage for a criticism of the relationalist presumption that one can read the relational logical form of belief sentences back onto the attributed belief states themselves.' (Matthews 2007, 102).

This could perhaps lead someone into believing that the relationalist view is no more than a relic from the past, but this diagnosis would be dead wrong when the reception of Wittgenstein's views is taken into consideration. As we will see, an important number of theories proposed in the past few years to deal with the problems concerning avowals assume that Wittgenstein's rejection of descriptivism is confined to first-person mental states attributions. The aim of this paper is to show that Wittgenstein's anti-descriptivist approach to the meaning of mental states ascriptions is not restricted to this particular kind of first person mental states attributions –avowals.

2. The scope of Wittgenstein's anti-descriptivism

Wittgenstein's defense of the plurality of language games, the idea that we are usually wrong in considering that the sole purpose of language use is to describe a portion of reality, goes hand in hand with his arguments against the alleged "descriptive" –or "relational"– nature of certain sets of expressions. This raises no particular debate amongst Wittgenstein's commentators and profiteers. Quite a different attitude can be found when the scope of this critique to relationalism is at stake. On the one hand, a number of authors favor a restrictive position, according to which Wittgenstein's anti-descriptivism only affects a particular kind of first-person mental states ascriptions (Bar-On & Long 2001; Fogelin 1976, 188 ff.; Fogelin 1996, 44 ff.; Bar-On 2004, 228 ff.; Finkelstein 2003; assumed in Wright 1998). Wittgenstein's anti-descriptivism would be for them an explanatory mechanism to provide a take on the nuances of avowals (particularly on immunity to error through misidentification), a realm in which it this theoretical attitude can be safely identified with *expressivism*. On the other side of the fence, some classical commentators have defended that Wittgenstein's arguments against descriptivism have a wider reach, and that mental states ascriptions in general are to be considered under this non-descriptivist light (cfr. Hacker 1996, 5 and ff; Hacker 2005, 246; Glock 1996).

Even though both sides have good historical reasons, as well as some textual evidence from Wittgenstein's corpus, to back their claims, we think neither of them does complete justice to the reach of Wittgenstein's arguments against descriptivism. Unlike the first group, we agree with Hacker and Glock that Wittgenstein's anti-descriptivism cannot be confined to a particular kind of

first-person mental states ascriptions. Against the second group, we think that Wittgenstein's take on the meaning of mental states ascriptions is not exclusively motivated by his rejection of the superficial grammar of the accusatives in these ascriptions. Wittgenstein offers a wide variety of arguments against the relational nature of mental states ascriptions, and his expressivism can only be understood by building up on this complex set of arguments.

Just to be clear: we do not claim that Wittgenstein did not establish a well-founded difference between first-person and third-person mental state attributions, we simply do not think that this difference hinges on the descriptive nature of the latter. Neither is our purpose to show that according to Wittgenstein mental states cannot be *described*. As a matter of fact he punctually gives examples of this activity. It is just not the job of mental state ascriptions to describe mental states.

3. Wittgenstein's *extended* anti-descriptivism

Our strategy is *threefold*. We will *first* provide textual evidence to make it apparent that Wittgenstein's non-relational treatment of mental states ascriptions is not restricted to avowals. *Secondly*, we will analyze different arguments provided by Wittgenstein against the relational nature of mental states ascriptions. *Finally*, in the next section, we will develop some ideas to try and make sense of the claim that by using third-person mental states ascriptions we are not describing, but expressing mental states.

Firstly, many times in PI first-person attributions are preceded by questions in the second grammatical person, and it is difficult to defend that a non-descriptive first-person ascription suffices to answer a question containing a descriptive second person attribution. In PI 677, for example, Wittgenstein writes:

> '"When you were swearing just now, did you really mean it?" This is perhaps as much as to say: "Were you really angry?"—And the answer may be given as a result of introspection and is often some such thing as: "I didn't mean it very seriously", "I meant it half jokingly" and so on. There are differences of degree here. And one does indeed also say "I was half thinking of him when I said that."'

The question 'Were you really angry?' is posed in the second person. If Wittgenstein's anti-descriptivism was restricted to first-person attributions, then this kind of question would have to be interpreted as a relational question about a state of affairs, similar to questions concerning the relative positions of the objects in our office, for example. 'Were you angry?' would have to be taken on a par with questions like 'is my computer on under the table?'. Only the latter, nevertheless, can be answered by providing an accurate description of the space in front of us. How is it possible to imagine a *descriptive* question successfully answered in *non-descriptive* terms?

Besides, a whole range of non-relational attributions are systematically expressed using second-person ascriptions: normative statements (see, e. g. PI 190, PI 214, PI 231). If reporting how I obey an order is not describing a situation in which a certain event is a proper interpretation of a rule, since 'any course of action can be made out to accord with the rule' (PI 201), what reason could we have to suppose that reporting how somebody else obeys an order is *descriptive* in any sense? Rules, as expressed in the second person ('do not feed gremmlins after midnight') are not *descriptions* in the very same sense in which a report of my behavior in accord with a rule ('I didn't feed the gremmlin after midnight') is not a *description* of the way my action accords with a rule. Accordance with a rule is not a fact to be described, and thus normative reports, irrespectively of the grammatical person that they use, are not descriptions either. The connection between the non-descriptive nature of mental state attributions and normative statements is made apparent, for example, in §PG 47: "When we say that somebody understands, we do not mean that he has a certain definition or image present to his mind. That is "mere history" ."Must he have such a picture present whenever we would say he was using the word "rouge" with understanding? (Think of the order: "Imagine a red patch")",

More importantly, Wittgenstein directly addresses third person attributions in a number of places (e. g. PI § 321, PI §402, PI §453, PI § 572, PI 573, etc.). Take the following excerpt from PI 453, for example: 'To say of an expectant person that he perceives his expectation instead of saying that he expects, would be an idiotic distortion of the expression'. Throughout PI, Wittgenstein does not avoid the attribution of mental states, like expectation in PI 453, in third person. There is no reason to suppose that his views on second and third-person attributions are going to be any different from first-person attributions, with respect to their *non-descriptive* nature. Moreover, his general characterization of intentionality is orthogonal to any distinction between first and third person mental states attributions (see PG pp 161-162; Zettel §55).

Secondly, Wittgenstein provides a series of arguments against the descriptive nature of mental states attributions that, together with those taken into account by Hacker and Glock concerning the confusion between two different kinds of accusatives, apply to first, second, and third person attributions alike. We will focus on Wittgenstein's rejection of the idea that it is always meaningful to ask about the duration of a mental state (PI §638-640 on intention, PI §llxi on 'struck by', PG §12 on understanding, Z §78 on hope, Z §286 on orders, etc.). Being descriptions distributions in space and time, it should be meaningful to ask about described event 'when did it happen?', 'for how long did it happen?', etc. Wittgenstein shows that these questions make no point when asked after some attributions of mental states.

> PI 661: 'I remember having meant him. Am I remembering a process or state?- When did it begin, what was its course; etc.?'
> PG 12: 'In order to get clearer about the grammar of the word "understand", let's ask: when do we understand a sentence? - When we've uttered the whole of it? Or while uttering it?'
> Z 78. 'Is "I hope..." a description of a state of mind? A state of mind has duration. So "I have been hoping for the whole day" is such a description; but suppose I say to someone: "I hope you come"--what if he asks me "For how long have you been hoping that?" Is the answer "For as long as I've been saying so"? Supposing I had some answer or other to that question, would it not be quite irrelevant to the purpose of the words "I hope you'll come"?'

It seems quite straightforward that the reasons why an utterance of 'I understood the sentence at 12:00 sharp' is meaningless should apply to 'you understood the sentence at 12:00 sharp', and 'he understood the sentence at 12:00 sharp' as well.

4. Wittgenstein on the expression of other people's mental states

Perhaps the most challenging passage for the view here presented is the following one:

> Remarks on the Philosophy of Psychology vol I, §63: 'Psychological verbs characterized by the fact that the third person of the present is to be verified by observation, the first person not. Sentences in the third person of the present: information. In the first person present: expression. ((Not quite right.)) The first person of the present akin to an expression.'

Is it possible to refrain from the idea that Wittgenstein is establishing here a principled distinction between first and third-person mental state attributions? Isn't *observing* unavoidably linked to *description*? We don't think it is.

Wittgenstein does not deny the existence of mental states, and his position doesn't preclude the possibility of observing or describing them. We can train somebody to *observe* mental states, to 'put himself in a favorable position to receive certain impressions' (PI ix), and emit a certain sound when they struck him. This would still not be *describing* them. We can nevertheless describe somebody's mental state in the following way: 'he wasn't feeling great this morning, he wondered what was the purpose of getting up before he even opened his eyes, he looked at the coffee machine for 5 minutes before he finally made some coffee, all that while his kids were waiting for breakfast, worried that they were going to be late for school'. We are describing a mental state, and we get closer and closer as we provide more details on the situation. We might even reach the point where no word –'abulia', 'depression', or any other– could be used to express his situation (cfr. PI 588).

How can we be said to express, as opposed to describe, somebody else's mental states? The problem becomes apparent once we realize that Wittgenstein's story about how words about sensations are learned is not easily exported outside the real of first-person attributions.

> 'This question is the same as: how does a human being learn the meaning of the names of sensations?-of the word "pain" for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour. '(PI 244)

How can a human being learn to substitute 'natural expressions of sensations' by words? How can one "express" somebody else's fear? How can a third-person ascription "manifest" somebody else's expectations, beliefs, desires, etc.? How can I substitute with an expression of mine a third person's mental state's ascription, the expressive behavior that the ascribee would have undertaken, had she been in the appropriate circumstances?

An answer to these worries –compatible with the reach that Wittgenstein concedes to his considerations on the meaning of mental states attributions– could be developed along the following lines. When I make a second or third-person mental state ascription, I'm displaying the attitude of a *spoke-person*. I'm making it apparent that the ascribee is in a position in which it could have made perfect sense for her to express herself via the use of the corresponding first-person ascription. For a moment I become his mouthpiece, so to speak, with the purpose of conveying information about about the circumstances she is in to others (or herself). One of the crucial improvements introduced by the invention of language, Hobbes explained, is the ability to *personate* other peoples's desires and interests (cfr. *Leviathan* 16.1, cfr. Pettit 2008, 70 and ff.). I can, in this sense, *represent* other people's mental states without entertaining a *representation* (viz. a description) of them. This sense of 'representation' can be used to understand, for example, how is it that I can be said to express my neighbor's anger.

Like communication agencies, third person attribution of mental states is not necessarily constrained by *acceptance*. Agencies' formal statements are not usually written exactly as the companies, families or individuals, would have done them. That is one of the main reasons why these services are required. For a statement to represent my interests at a certain point, my explicit understanding or acceptance of it is not required. In the same vein, when we ascribe mental states to others, we do not necessarily have an obligation to use the words the ascribee would have used to manifest her situation.

It makes no sense, according to this view, to check the world in order to find the state of affairs that would make a mental state ascription true or false. Mental state ascriptions are not claims about how the world is. We learn to master these ascriptions when we are proficient in our grasp of the circumstances that would make such an ascription appropriate. It is not only to ourselves that we can give voice to by ascribing mental states, we can also act as a spoke-person for others, explaining the circumstances they found themselves in, and predicting their behavior. This can be cashed out in terms of *expressing* mental states, rather than *describing* them.

REFERENCES

Bar-On, D. and Long, D.C. 2001. "Avowals and First-Person Privilege," *Philosophy and Phenomenological Research* 62: 311-335.
Bar-On, D 2004. *Speaking My Mind*. NY: Oxford University Press.
Finkelstein, D. 2003. *Expression and the Inner*. Cambridge, Mass: Harvard University Press.
Fogelin, R.J. 1976. *Wittgenstein*. London: Routledge.
Hacker, P (2005). "Of Knowledge and of Knowing that Someone is in Pain." In Pichler and Saatela (eds) *Wittgenstein: The Philosopher and his Works*. Bergen: The Wittgenstein Archives and the University of Bergen.
Pettit, P (2008): *Made with Words. Hobbes on Language, Mind, and Politics*. Princeton: Princeton University Press
Wittgenstein, L.
1961/1921. *Tractatus Logico-Philosophicus*. London: Routledge.
1958/1953. *Philosophical Investigations*. Oxford: Blackwell. 1965/1929. "A Lecture on Ethics." Philosophical Review 74: 3-12.
1967. *Zettel*. Oxford: Blackwell.
1980. *Remarks on the Philosophy of Psychology*. Oxford: Blackwell.
Wright, C. 1998. "Self-Knowledge: The Wittgensteinian Legacy." In *Knowing Our Own Minds*. Wright, C, Smith B., and MacDonald, C. eds. Oxford: Oxford University Press.

UNIDAD CONCEPTUAL VERSUS MULTIPLICIDAD DE GRUPOS DE DEFERENCIA

SILVIA ANDRÉS BALSERA
Universidad Complutense de Madrid
s.andres@filos.ucm.es

ABSTRACT: It is proposed an exploration of the tension existent about the compatibilization the 'multiple deference' phenomenon, an extension of the division of linguistic labour hypothesis developed by Putnam, with an externalist framework about the individuation of our concepts.

KEY WORDS: deference, division of linguistic labour, externalism, conceptual competence

Nuestro propósito en esta comunicación es explorar la posible compatibilidad de estas dos tesis. En primer lugar, la tesis de que existe un repertorio conceptual compartido por los hablantes de una comunidad lingüística y que dicho repertorio constituye la condición de posibilidad del entendimiento mutuo entre hablantes. En segundo lugar, la tesis de que, en ocasiones, dentro de una misma comunidad, existen varios grupos de expertos que establecen, en relación a un mismo término, condiciones de aplicación relativamente distintas. Proponemos disminuir la tensión entre ambas ideas a partir de una visión de las condiciones de posesión conceptual cercana a la del externismo (social) de Burge.

Primeramente, convendría tomar en consideración el contenido de cada una de las tesis enunciadas. En lo que se refiere a la primera de ellas, mantenemos que se trata de una tesis que intuitivamente goza de una gran plausibilidad. En tanto que los hablantes de una comunidad lingüística utilizan un lenguaje común con fines comunicativos están interesados en entenderse usando las mismas oraciones (asertivamente) para expresar, generalmente, los mismos pensamientos. La existencia de un repertorio conceptual compartido y estable constituye la condición de posibilidad de la comunicación entre hablantes de un mismo lenguaje.

En cuanto la segunda tesis, ésta toma pie en la hipótesis de la *división social del trabajo lingüístico* presentada por Putnam en (1975), aunque se requieren ciertas cualificaciones sobre el

planteamiento inicial de dicha hipótesis para obtener la formulación que hemos presentado. De acuerdo con dicho planteamiento, los términos de nuestro lenguaje (especialmente los términos de género natural) pueden tener condiciones de aplicación que sólo dominan un subconjunto de hablantes de nuestra comunidad o "expertos"; típicamente, el grupo de científicos conocedor de las propiedades subyacentes de los géneros naturales. Sin embargo, para Putnam este hecho no impide que la mayoría de los hablantes puedan usar los términos de su lenguaje aceptablemente bien, en la medida en que estos hablantes "profanos" deleguen la tarea de conocer en detalle las condiciones de aplicación de los mismos al subconjunto de expertos relevante. Ahora bien, pensamos que la formulación inicial de esta hipótesis es un tanto restringida, debido, quizá, a que su uso está subordinado a la defensa de la semántica de los términos de género natural (y, en menor medida, de los términos de género artificial) que Putnam presenta en (1975). Dicha subordinación tiene la siguiente consecuencia. Si la semántica de los términos de género natural de Putnam es correcta y la pertenencia a la extensión de un género viene determinada por la posesión de determinadas propiedades subyacentes por parte de las instancias del género, sólo habrá un grupo de expertos en relación al término de género natural relevante, a saber, aquel grupo de hablantes que conoce las propiedades subyacentes y, con ellas, la condición determinante para la aplicación del término. Sin embargo, pensamos que es plausible mantener que la división del trabajo lingüístico es un fenómeno ubicuo, que se extiende más allá de los términos de género natural (como poco, también a los nombres propios[1] y otros términos generales como los términos de género socio-legal). Además, en relación a muchos términos no existe una distinción tajante entre expertos y profanos, sino que, más bien, la competencia de los hablantes en el dominio de un término se distribuirá de manera gradual en el seno de la comunidad lingüística. Por último y sirviéndonos de los ejemplos que presentaremos en lo siguiente, mantenemos que, en ocasiones, con respecto a un mismo término no existe un único grupo de expertos reconocido como tal por los hablantes de una comunidad, según se desprende de las tesis de Putnam, sino que existirán distintos grupos de expertos que pueden determinar condiciones de aplicación para el término relativamente distintas entre sí.

Una vez descritas las tesis a favor de cuya compatibilidad abogamos, cabe explicar la tensión que, *prima facie*, existiría entre ambas. Así, en primer lugar hemos mantenido que con respecto a un término lingüístico pueden existir múltiples grupos de deferencia, esto es, conjuntos de hablantes reconocidos socialmente como expertos en su dominio. A modo de ejemplo, es

[1] Así, Strawson, en (1979), p. 182, mantiene que las descripciones definidas que un hablante asocia con un nombre propio y que, para este autor, sirven para determinar la referencia del nombre, son descripciones que "pueden incluir la referencia de otro [hablante] a tal particular". Consúltese también al respecto el empleo que hace Searle en (1983), pp. 243, 245, 248, de la expresión "descripciones parásitas".

plausible pensar que la práctica totalidad de los hablantes del castellano confían por defecto en los Académicos de la Lengua para determinar las condiciones de aplicación de los términos de su lenguaje, incluido el término "matrimonio". Asimismo, un subconjunto de hablantes del castellano confía las condiciones de aplicación de ese mismo término a los sacerdotes católicos. Pero, además, dentro de nuestra comunidad lingüística hay otros grupos de individuos que no reconocen a los sacerdotes como expertos en este sentido y confían más bien en los jueces, expertos que sancionan y disuelven los matrimonios. El problema es que estos tres grupos de expertos (pues así los reconocen los hablantes del castellano) determinan condiciones un tanto diferentes para la aplicación del concepto relevante, ya que los dos primeros grupos no autorizan la aplicación del concepto a uniones homosexuales y el segundo, además, tampoco autoriza la aplicación de dicho concepto a la unión de dos personas divorciadas. A partir de esta situación, debemos preguntarnos si el hecho de que existan grupos de expertos "rivales" dentro de una misma comunidad, que establecen distintas condiciones de aplicación para un término, amenaza la identidad de los conceptos compartidos comunitariamente. Si ocurriera que estos hablantes expresan distintos conceptos usando un mismo término (en función de cuál sea el grupo de expertos en el cual confían), tendríamos que admitir que no existe tal cosa como un concepto compartido relevante, sino que, más bien, el hecho de que los hablantes de una comunidad empleen el mismo término (o, al menos, términos homófonos) no constituye garantía alguna de que estén expresando los mismos conceptos.

No obstante, nuestra tesis consiste en que la relativa unidad y estabilidad del repertorio conceptual comunitario puede mantenerse,[2] a pesar de la existencia de múltiples grupos de deferencia, si adoptamos una comprensión de la posesión conceptual como la propuesta por Burge en (1986) y (1989) y apelamos, además, a la noción wittgensteniana de "parecido de familia" como un criterio débil para establecer la identidad del repertorio conceptual de los hablantes.

En lo que se refiere a lo primero, suscribimos la tesis, típicamente externista, de que un individuo puede poseer un concepto aunque no domine por completo sus condiciones de aplicación, siempre y cuando se encuentre en las relaciones apropiadas con su entorno (social y natural). En consecuencia, no será preciso que un hablante individual conozca *todas* las condiciones de aplicación para un término para ser competente en el uso del término relevante y el concepto expresado por él. Además, hemos modificado la formulación inicial de la hipótesis de la división del trabajo lingüístico para rechazar que, por lo general, con respecto a un término dado, exista una

[2] Naturalmente, pueden encontrarse casos en los que un subconjunto de hablantes toma la decisión de otorgar un nuevo significado a una determinada expresión lingüística. Este fenómeno, distinto del que discutimos, suele producirse en el contexto de creación de jergas.

división en el seno de una comunidad de hablantes entre expertos y profanos, abogando, más bien, por la idea de que la competencia de los hablantes en el dominio de *un mismo concepto* es gradual (Burge 1986, pp. 258 y ss.).

En segundo lugar, esta idea puede reforzarse a partir del siguiente criterio de identidad conceptual, a saber, si entre los criterios de aplicación determinados con respecto a un término "T" por los distintos grupos de deferencia dentro de una misma comunidad existe una relación de parecido de familia, entonces consideraremos que dichos grupos son "rivales" con respecto a un *mismo* concepto. Utilizando el ejemplo anterior, según un conjunto de expertos, el concepto de matrimonio se aplica a uniones de dos personas de distinto sexo, mientras que otro grupo aceptará las uniones polígamas como matrimonio, otros rechazarán la poligamia pero no mantendrán el requisito de la diferencia sexual, etc. La cuestión es que entre todas estas concepciones de un mismo concepto hay un cierto *parecido de familia* (Wittgenstein 1953), suficiente para admitir que todos ellos disputan sobre las condiciones de aplicación del mismo concepto.

Para terminar, ofrecemos tres argumentos adicionales para sustentar nuestra propuesta. En primer lugar, pensamos que el hecho de contemplar la existencia de múltiples grupos de deferencia como una amenaza a la identidad conceptual toma pie en una visión de la multiplicidad deferencial similar a la del ejemplo analizado (el término "matrimonio"), es decir, como si distintos grupos de hablantes profanos dentro de una misma sociedad confiaran, cada uno de ellos, en diferentes expertos "rivales". Sin embargo, en ocasiones es un *mismo* individuo el que confía, con respecto a un concepto dado, en distintos expertos. Así, un mismo hablante puede delegar las condiciones de aplicación del término "cáncer" tanto en los médicos al uso como en las practicantes de medicinas alternativas (de hecho, hay médicos al uso que participan en dichas prácticas), aun cuando las condiciones de aplicación de ese término según este último grupo involucren conceptos como el de *chakra* que no son contemplados por el primero. Ante un caso como éste es más difícil establecer una diferencia en el concepto, ya que supondría contemplar a este hablante en una situación de "esquizofrenia conceptual".

En segundo lugar, el hecho de que consideremos que existen genuinos desacuerdos entre hablantes que confían en distintos conjuntos de expertos respalda la tesis de la unidad conceptual de base. De este modo, si el concepto expresado mediante el término "T" por cada uno de los hablantes fuera distinto, la proposición expresada por la oración "T es P", cuya verdad uno afirma y otro niega, no sería la misma, en virtud de que involucra distintos conceptos, ni existiría un desacuerdo genuino, sino una mera disputa que involucra un término equívoco. Desde luego, nuestra tesis no conlleva mantener la unidad conceptual a toda costa y negar testarudamente que existan discusiones

meramente verbales. Sin embargo, dada la ubicuidad del fenómeno de la deferencia múltiple, si no estableciéramos una unidad conceptual comunitaria de base a pesar de dicho fenómeno las disputas verbales serían la norma y no la excepción.

En tercer lugar, cabría alegar que existen evidencias de que los propios términos de género natural, que, en un principio, constituyen una clase de expresiones que encaja relativamente bien en las coordenadas iniciales de la hipótesis de la división del trabajo lingüístico, también están sometidos al fenómeno de la deferencia múltiple. A este respecto, aun cuando parece que los científicos constituyen, si no el único grupo de expertos en relación a estos términos, sí el grupo de deferencia privilegiado, en ocasiones no ocurre de este modo. El motivo es que la denominación y distinción de géneros naturales en contextos ordinarios se rige por intereses muy variados que, en ocasiones, atienden a ciertos rasgos de las entidades naturales que, desde el punto de vista científico, carecen de relevancia. A modo de ejemplo, el interés sociológico, el asombro intelectual que suscita una clase de animales o plantas o la familiaridad que se establece con ellos, la importancia económica de estas entidades, etc., son factores que determinan la introducción en el lenguaje y el uso ordinario de términos de género natural. En consecuencia, de acuerdo con el fin que los hablantes persigan en el trato con las entidades naturales, pueden variar los criterios de individuación de los géneros naturales y, en función de ello, el grupo de deferencia para el uso del término de género natural relevante. En palabras de Hanfling, "[e]l científico puede conocer, *qua* científico, que el anillo dorado que él compró en la joyería no es 'realmente' de oro, sin negar que sea realmente de oro en el sentido del joyero" (2000, p. 229).

REFERENCIAS

Burge, T. 1986, "Intellectual Norms and Foundations of Mind", *Burge* (2007), 254-274.
Burge, T. 1989, "Wherein is Language Social", *Burge* (2007), 275-290.
Burge, T. 2007, *Foundations of Mind*, Nueva York, Oxford University Press.
Hanfling, O. 2000), *Philosophy and Ordinary Language: The Bent and Genious of our Tonge*, Nueva York/Londres, Routledge.
Pessin, A. y S. Goldberg (eds.) 1996, *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The meaning of 'meaning'"*, Londres, Sharpe.
Putnam, H. 1975, "The Meaning of 'Meaning'", *Pessin/Goldberg* (1996), 3-52.
Searle, J. R. 1983, *Intentionality: An Essay in the Philosophy of Mind*, Oxford, Cambridge University Press.
Strawson, P. F. 1979, *Individuals: an Essay in Descriptive Metaphysics*, Londres, Methuen.
Wittgenstein, L. 1953, *Philosophical Investigations*, Oxford, Blackwell.

THE SINGULAR THOUGHT STRATEGY AND THE CONTENT OF PERCEPTION

MARC ARTIGA

ABSTRACT: It is standardly assumed that some of our thoughts are singular and that in some cases these singular thoughts are grounded in experience. Recently, some people have argued that these two claims show that the content of experiences should also be singular. This is what I call the 'Singular Thought Strategy'. In the paper I argue that the Singular Thought Strategy is flawed. I show that the fact that we have singular thoughts and that they are grounded in experience does not support the claim that the contents of experience are singular.

KEYWORDS: Perception, Experience, Representational Content, Singular Thought, Grounding

One of two most controversial topics in current philosophy of perception is whether perceptual experiences have representational content and, if so, how should be conceived them. In this paper I will focus on the second question and for this reason I will be assuming that experiences represent the world as being in a certain way.

Among those who think that experiences have representational content, there are two main ways of cashing these contents out. On the one hand, the standard view claims that perceptual states are endowed with *singular* content, that is, that the content of perceptual states involves the object of perception (Bach, 2007; Burge, 2010; Tye, 2009; Schellenberg, 2010, 2011; Siegel, 2011). According to this approach, which I will call the 'Singularity View', if at t_1 John is perceiving object$_1$, then the accuracy conditions of his perceptual state involves object$_1$. Similarly, if at t_2 object$_1$ is changed by a qualitatively identical but numerically distinct object$_2$, then the content of his perceptual state is different; at t_1 John's perceptual state had the content that object$_1$ has such and such properties and at t_2 it has the content that object$_2$ has such and such properties. This is said to be so even if the two experiences are indistinguishable from the first-person point of view. On this proposal, the contents of perception can be adequately described in English with the following expression: *that object has such and such properties.*

An alternative view, which following Tye (2009) I will call the 'Existential View', holds that the content of perceptual states is general, that is, that veridical experiences of qualitatively identical but numerically distinct object can share content (Davies, 1992). Hence, on the Existential View we

can appropriately describe the contents of perception by using an existential quantification as follows: *there is an object with such and such properties*.

In this paper I argue against one of the standard arguments in favor of the Singularity View. I will show that this argument, which I dub the 'Singular Thought Strategy', fails to support the Singularity View over the Existential View.

1. The Singular Thought Strategy

The argument focuses on certain cognitive roles perceptual experiences are supposed to play and argues that perception can play these roles only if the contents of perceptual experiences are singular.

In a nutshell, the argument is the following: the first premise claims that some of our *thoughts* are singular, that is, some of our thoughts are object-involving. The second premise of the argument says that some of these singular thoughts are grounded in perception. That is, we can have certain singular thoughts *in virtue of* being in perceptual contact with the object of belief. For instance, in certain situations my belief THIS BALL IS RED is a contentful state that involves a particular ball in virtue of the fact that I am perceiving a particular ball. Finally, in order to conclude the argument, we need the premise that a perceptual state grounds a belief only if the content of the believe is already contained in the content of the perceptual state. In other words, a perceptual state can ground my singular belief about P only if the content of the perceptual state already contains P. The conclusion of the argument is that the content of (at least some) perceptual states must be singular, in the sense that it has to be object-involving (Siegel, 2010; Campbell, 2002 p. 124). So this argument clearly lends support to the Singularity view over the Existential View.

This argument can be found, for instance, in Siegel (2009, p. 157):

> So far, we can conclude that experiences have singular contents. (...) This conclusion vindicates the idea that experiences play certain special roles vis a vis *de re* mental states by providing relations to particular objects and by attributing properties to them. For example, experiences make it possible for us to form de re mental states about the objects we see, and experiences provide evidence for beliefs about those particular objects. The fact that experiences have objectually singular contents may help explain how they play these roles.

I think the argument and the intuition underlying it has a very strong appeal, but it is utterly flawed. To see why, we need first to formulate the argument in more detail.

2. Grounding Thoughts in Perception

First of all, we must define the relevant notion of grounding. The notion that seems to be in place is the following:

> **Grounding:** A single experience E grounds the capacity to have a belief involving an object or property P iff a suitable concept-user who has E will have the capacity to have beliefs involving P because he has E.

Roughly, a suitable concept user is one who is not cognitively impaired in comparison with a normal human. The important point about this notion of grounding is that an experience grounds the capacity for having a belief about an entity P if the subject has the capacity to entertain thoughts involving P *because of* that experience.

An important remark about this definition of grounding is that it relies on the idea that two people can have the same experience E (type). Hence, at this point it might be useful to specify individuation conditions for experiences (that will also be important in the main argument). In what follows, I will assume that a (token) experience E and a (token) experience E' are type-identical iff (1) they have the same content and (2) they have the same phenomenal character. If E and E' differ in content or phenomenal qualities, we should say they are different experiences.

Having defined the key relation of grounding, we can spell out the grounding principle that constitutes a key premise of the argument:

> **Grounding Principle:** If an experience E grounds the capacity to have beliefs involving P (without imaginative extrapolation), then P enters into the content of experience E.

The grounding principle is the intuitive claim that lies behind Campbell and Siegel's argument for the singular content of perceptual states. The idea is that an experience grounds a thought involving a certain property only if that property is also represented in the experience. In other words: if the content of experience does not involve P, then the experience does not ground the capacity to have beliefs involving P.

I have excluded the use of imaginative extrapolation in the grounding principle in order to exclude cases where an experience causes one to have certain thoughts about an entity because some memories come to one's mind or similar cases (Pautz, 2009). The important point here is that the capacity to think about P must be somehow directly caused by the experience. The exclusion of imaginative extrapolation is not supposed to rule out the use of certain concepts one already possesses in constructing a thought. That would be absurd; since in any thought one attributes a certain property to an entity, thoughts usually involve a concept one already possesses (the predicate). The key issue here, is that the entity one is thinking about (the object referred to) must

be determined without imaginative extrapolation. I certainly admit the expression 'imaginative extrapolation' is a bit imprecise, but an extended discussion of that point would lead us too far from our present concern.

Now that we have a suitable 'grounding principle', we are in position to spell out the Singular Thought Strategy against the Existential View:

1. Any veridical experience E grounds the capacity to have a perceptual belief that involves a particular object (without imaginative extrapolation).

2. If a veridical experience E grounds the capacity to have beliefs involving P (without imaginative extrapolation), then P enters into the content of E (Grounding Principle).

C) Therefore, particular objects enter into the content of experiences (contents are object-involving).

The conclusion, thus, is that the Singularity View is true.

In order to resist the argument I have to either show that one of the premises is false or that the argument is invalid. Since the argument is clearly valid, only the first option is available. In particular, I will present two sorts of considerations that suggest premise 2 (the Grounding Principle) is probably false.

First, there are some counterexamples that plausibly show 2 to be false. Suppose Jack and John are twins grown up on Earth, who have never heard or been in contact with ethanol (whose chemical basis is CH3CH2OH). One day John sees ethanol in a glass and Jack sees a substance that looks exactly the same way but which happens to be XYZ (a different substance). In this case, it is extremely plausible that the perceptual content of both John and Jack is the same. Surely, John's perceptual content when first confronted with ethanol does not involve CH3CH2OH. Nevertheless, John could have had the following true demonstrative thought about this substance: THIS IS TOXIC. The content of this thought surely involves CH3CH2OH and predicates of it that it is toxic. In contrast, Jack's demonstrative thought THIS IS TOXIC refers to XYZ. This is a clear case where an experience grounds a thought without the content of the thought being contained in the content of the experience.

Of course, one could insist that John and Jack's experiences have different contents and that this fact explains why they ground different thoughts. But notice the supporter of the grounding principle is committed to the claim that *the first time* John is confronted with ethanol, his experience

already represents there being CH3CH2OH, since this experience already grounds the capacity to have thoughts about CH3CH2OH. I think that even those who accept that our experiences have a rich content would hardly accept this consequence. I think this view puts some pressure on the Grounding Principe.

Secondly, I think there is a more general worry with the Grounding Principle. Experiences ground thoughts about many different entities. Whenever I have an experience, my experience grounds the capacity to have singular thoughts about a variety of entities: I can demonstratively think about that individual entity (for instance, when I truly think THIS IS MY CAT), about a natural kind (when I truly think THIS IS A CAT), about a bunch of molecules, (when I truly think THIS IS SET OF MOLECULES ARRANGED IN A CERTAIN WAY) and so on. Notice that in all these examples the referent of the demonstrative element (the referent of 'this') should be identified before anything is predicated of it; after all one can mistakenly refer to a natural kind and attribute to it certain properties that only molecules have. The fact that we possess the capacity to refer to all these different kinds of entities generates an important problem for the grounding principle: if the grounding principle were right and the fact that a perceptual state grounds the capacity to entertain these thoughts were sufficient for the perceptual experience to be representing these properties, then we would arrive at the unattractive conclusion that in perception we always represent natural kinds, individuals, microphysical entities,... That is an extremely unappealing consequence of the theory. So, the Grounding Principle should not guide our theories on perceptual content.

So I think the Grounding Principle is mistaken. However, one might reasonably ask: if this principle has clear counterexamples and such unappealing consequences, why so many people think it is plausible? A full rebuttal of the Singular Thought Strategy should explain why people find this reasoning convincing. What I will argue next is that Grounding Principle is intuitively compelling because there is a very similar principle that is probably right. However, I will also show that while this similar principle is right, it cannot provide an argument in favor of the Singularity View. Let me first present the principle and discuss how it bears on the argument about the contents of perception.

3. Strong Grounding Principle

As we saw, John's experience grounds his capacity to entertain thoughts about CH3CH2OH and Jack's experiences grounds his capacity to have thoughts about XYZ. A plausible explanation of the fact that John and Jack's experience grounds the capacity for different thoughts is that CH3CH2OH *happens to be* the colorless liquid John is confronted with and XYZ is the substance

that *happens to be* the colorless liquid that Jack is confronted with. In contrast, think about properties that are very probably represented in experience, such as *being red*. In this case, there seems to be a *necessary* connection between having these experiences and having a capacity to entertain beliefs with these contents (in normal subjects). It seems that if a human perceives red things, she would necessarily have the capacity to entertain red-involving thoughts. So one might think that if we formulate the Grounding principle using a stronger notion of grounding (necessary grounding) this sort of counterexamples will be avoided. Let us see whether this intuition is right.

So, let us consider how the argument presented above would look like if we used the stronger relation of necessary grounding between the experience E and a content involving an object P. That is:

> **Necessary Grounding**: A single experience E necessarily grounds the capacity to have a belief involving P iff it is necessary that suitable concept-users who have E will have the capacity to have beliefs involving P because they have E.

For instance, an experience as of a red object necessarily grounds the capacity to have beliefs involving being red, since it is necessary that any normal human has the capacity to have beliefs involving redness in virtue of representing red in experience. In contrast, an experience as of a red object does not necessary ground a belief involving tomatos, because it is not necessary that a normal human has the capacity to have beliefs about tomatoes in virtue of having this experience (there are many possible worlds where there are no tomatoes).

An interesting question here is whether the necessity involved here is nomological, metaphysical or logical necessity. Since my point can be made with either of these modal notions, I will not try to argue in favor of any of them.

Using the notion of necessary grounding, we can derive the key premise of the new version of the Singular Thought Strategy:

> **Strong Grounding Principle**: If an experience E necessarily grounds the capacity to have beliefs involving P (without imaginative extrapolation), then P enters into the content of E. (Pautz, 2008, p. 18)

Now, I think the Strong Grounding principle is very plausible. If an experience E necessarily makes it possible for any concept-user to have beliefs involving P, then it seems that the experience should itself represent P.

Now, the supporter of singular contents might seek to formulate the argument using this stronger notion of grounding:

1. Any veridical experience E necessarily grounds a demonstrative thought that involves particular objects (without imaginative extrapolation).

2. If an experience E necessarily grounds the capacity to have beliefs involving P (without imaginative extrapolation), then P enters into the content of E.

C) Therefore, particular objects enter into the content of experiences.

The argument is clearly valid and I have already granted 2; in this occasion, however, the problem lies in premise 1. Why should we think that experiences necessarily ground demonstrative thoughts involving particular objects? that is, why should we think that any normal human that has the same experience as I do will have the capacity to develop demonstrative thoughts about the same object? Prima facie, premise 1 seems wrong. If I am seeing a tomato, I can develop demonstrative thoughts about tomatoes, but there are many scenarios (in the actual world and in nearby possible worlds) where someone with the same experience fails to develop the thoughts about tomatoes.

The crucial problem with premise 1 of the second version of the argument is that the only way of lending support to it is by assuming the conclusion. For any experience E, E necessarily grounds a demonstrative thought about a particular object o only if the content of E involves o. Suppose we think the content of experience has the form of an existentially quantified content as suggested by supporters of the Existential View. According to this approach, the content of perceptual experience can be adequately expressed as *there is an object with such and such properties.* If this were so, then my experience E would not necessarily ground a demonstrative thought about a particular object o, because E could have been made true by many different objects (and thus, it could enable the creation of demonstrative thoughts concerning many different objects). In other words, if we assume the Existential View, for any experience E there might be no object, such that the experience E necessarily grounds beliefs about that object, because there are possible situations where a competent and normal user has the same experience E but she has not the capacity to develop beliefs about that particular object o. On the other hand, if the Singularity View is right and the content of perceptual experiences is object-involving (i.e. it can be expressed in the following way: *that object has such and such properties*), then premise 1 is right. So premise 1 of the argument is true only if we assume that objects (among others) figure in the contents of the

experiences. However, this the conclusion of the argument, so it is precisely what we were suppose to show. Therefore, the argument is only compelling if the conclusion is already assumed.

Conclusion

The conclusion of this discussion is that the fact that our experiences ground the capacity to entertain thoughts about particular objects does not constitute an argument for preferring the Singularity View over the Existential View. On the one hand, the Grounding Principle is probably false, so the Singular Thought Strategy does not go through. On the other, the Strong Grounding Principle seems to be right but fails to provide a motivation for the Singularity View. Therefore, the fact that our experiences ground the capacity for entertaining certain singular thought cannot be used then against the Existential View. The Singular Thought Strategy has been blocked.

REFERENCES

Bach, K., 2007, 'Searle against the world: how can experiences find their objects?', Tsohatzidis S. (ed.) *John Searle's Philosophy of Language*, Cambridge University Press, Cambridge
Burge, T, 2010, *The Origins of Objectivity*, Oxford, OUP.
Davies, M., 1992, 'Perceptual Content and Local Supervenience', *Proceedings of the Aristotelian Society,* 92, 21–45.
Pautz, A., 2009,' What are the Contents of Experience?', *Philosophical Quarterly,* 59 (236):483-507.
Siegel, S., 2011, *The Contents of Visual Experience,* Oxford, OUP
Schellenberg, S., 2010, 'The Particularity and Phenomenology of Perceptual Experience'. *Philosophical Studies*, 149: 19-48
Tye M. 2009, *Consciousness Revisited: Materialism without Phenomenal Concepts,* Cambridge MA, MIT Press.

EL MODELO ANIMAT EN LA INTELIGENCIA ARTIFICIAL BIO-INSPIRADA

A. BENÍTEZ, E. G. BUENDÍA
U. Complutense de Madrid, U. Complutense de Madrid

ABSTRACT: This paper presents the model Animat. First sets the positions of Wilson and Beer. The following summarizes these positions in a schema. Schema to be followed by any researcher who wants to work with this model. The work ends with the presentation of the basic outline of several experiments in the model.

Key words: Artificial Life, Bio-Inspired Artificial Intelligence, The Animat Model.

1 El modelo Animat

El término "animat" apareció por vez primera en un artículo de Stewart W. Wilson de 1985. Con dicho término Wilson designa un autómata o agente dinámico que habita en un entorno artificial. Después de los trabajos del propio Wilson, de los de Beer y Brooks, entre otros, hoy designa un modelo computacional de experimentación.

Los trabajos de Wilson y Beer coinciden en un presupuesto básico, nuclear, a saber: suponen que la inteligencia es un mecanismo de adaptación con que nacen dotados los individuos y sin el cual la especie correspondiente no perviviría.

Lo que hemos llamado "modelo Animat" es un planteamiento que inserta la investigación de la inteligencia artificial en la línea de la adaptación biológica. Por otra parte, el modelo Animat es eminentemente experimental. Por ello, el desarrollo de las investigaciones se logra mediante experimentos computacionales o robóticos.

1.1 Wilson

En el primer trabajo de 1985 Animat pretende ser como un animal cuyo medio es una matriz de datos de límites periódicos, es decir, un espacio computacional de forma toroidal. Cada elemento de la matriz de datos puede contener un objeto. Son objetos: obstáculos (T), comida (F), espacios vacíos (blancos) y el propio animat (*). En el resumen de este artículo, Wilson señala con claridad los

rasgos principales del modelo Animat. El primero es que se trata de un experimento computacional. El diseño, segundo, incluye dos elementos complementarios: un autómata ("artificial animal" lo llama Wilson) y un entorno. El tercer rasgo es que el animal artificial actúa en el entorno, por ello, es un autómata con dinamismo propio. El último es que el experimento tiene el objetivo de medir el grado de adaptación del animal artificial. El experimento concreto que Wilson explica incluye un rasgo más: comprender qué sea la inteligencia.

Para conseguir este objetivo, Wilson hace suya una definición de inteligencia dada por van Heerden:

> but the best definition of intelligence we have found is the following, from the physicist van Heerden:
>
> > Intelligent behavior is to be repeatedly successful in satisfying one's psychological needs in diverse, observably different, situations on the basis of past experience.

de la que el propio Wilson peralta los siguientes aspectos: el ente estudiado ha de poder encontrarse en diferentes situaciones posibles; en cada situación, dicho ente ha de tener éxito cara a satisfacer alguna necesidad constitutiva; al afrontar una situación, ha de basarse en experiencias propias obtenidas en el pasado.

La definición de van Heerden implica, a nuestro juicio, una estructura constitutiva del agente cuyos elementos se pueden enumerar así:

1. ha de contar con algún receptor sensitivo
2. y con algún efector
3. igualmente con una memoria de lo sentido
4. además, con la capacidad de generalizar y formar representaciones universales a partir de lo recordado
5. y con una serie de reglas generales que ligan representaciones universales con algún efector.

Por ello, Wilson adoptó la idea de un Clasificador Genético (Holland) para generar nuevas reglas generales mejores o más adaptadas que las iniciales.[1]

[1]No podemos extendernos aquí en una explicación pormenorizada del trabajo mencionado. A. Benítez lo hace en el tercer tomo de sus *Fundamentos de I. A.* que aparecerá próximamente.

En un segundo trabajo de 1991, Wilson hizo un presentación general del modelo experimental. Lo que resulta más claro del trabajo es, a nuestro juicio, la división del modelo en dos componentes: entorno y animat. Es decir, que el investigador que se acoja a este modelo ha de describir siempre cuáles son las características con las que ha diseñado el entorno y cuál es la estructura con que ha dotado a animat.

1.2 Beer

En el libro *Intelligence as adaptive behavior: an experiment in computational neuroethology*. Academic Press, Boston, 1990, Beer afirma que su libro tiene dos objetivos fundamentales:

- El primero consiste en dar argumentos a favor de la idea de que la inteligencia humana es un mecanismo adaptativo.
- El segundo, presentar una nueva metodología experimental para construir agentes autónomos.

El primer objetivo del libro, defender la idea de que la inteligencia es un mecanismo adaptativo —quizá no el único que el hombre posee—, se opone a otra idea de inteligencia. Según Beer, esta otra idea de inteligencia ha subrayado las habilidades mentales del hombre que parecen diferenciarle: pensamiento lógico, lenguaje, acción racional, etc. Cabe decir que los defensores de esta idea de inteligencia son el "adversario" de Beer.

Beer cree que la idea de inteligencia como solución de problemas, razonamiento lógico, acción moral, acción planeada, etc., ha predominado hasta la fecha en que escribe el libro que comento. Los nuevos trabajos en robótica (Brooks), los nuevos planteamientos en experimentos computacionales con agentes (Wilson), el desarrollo de otras líneas de investigación en inteligencia artificial bio-inspirada, han producido un giro en las investigaciones bajo el supuesto general de que la inteligencia ha de ser contemplada a la luz de los mecanismos de adaptación. Para Beer, la inteligencia artificial clásica, al aislar los procesos mentales de razonamiento, lenguaje, toma de decisiones, etc., estableció una disociación entre "mundo" y "mente". Gracias a su mente, el hombre se hace "representaciones" del mundo, y gracias a estas representaciones, es capaz de actuar de vuelta sobre el mundo. El mundo aparece así, primero, como independiente del hombre y como algo que este ha de comprender. Tras cierto esfuerzo, el hombre "comprende" el mundo, es decir, se hace una cierta representación del mundo. Combinando estas representaciones, el hombre elabora sus planes de actuación en el mundo. A esta concepción llama Beer la hipótesis nuclear del programa de investigación en inteligencia artificial, hipótesis que se extendió a las ciencias cognitivas.

Esa hipótesis nuclear ha animado las investigaciones en inteligencia artificial durante décadas. Pero si nos volvemos hacia los resultados de ese trabajo de investigación, ¿qué tenemos? Afirma Beer que son técnicas cuya aplicación requiere un campo perfectamente delimitado y normalmente angosto (*narrow*). En cuanto se intenta superar esos límites y se intenta aplicarlas a problemas cotidianos, las técnicas se muestran en extremo frágiles (*brittle*).

A la primera pregunta, qué significa ser inteligente, contesta positivamente Beer diciendo: inteligencia es un mecanismo que permite conductas adaptadas a un mundo cambiante, complejo, impredecible. Por "conducta adaptada" se entiende una que permita la supervivencia del animal que la lleve a cabo. Los mecanismos gracias a los cuales esa conducta se desencadena, es posible, serían, por tanto, una característica adaptativa. Esta idea parece razonable. Sin embargo, hay que pensar casos como el siguiente: a veces, la carrera veloz del antílope le permite escapar del león depredador; otras no, y perece. ¿Cabe pensar que han cambiado los mecanismos puestos en juego de una vez a otra? La pregunta tiene la intención de subrayar que la idea de adaptación no lleva consigo la de éxito. Y si se cae en esta cuenta, se observará que la adaptación dice relación más a la especie que al individuo. Podemos aceptar que los antílopes (especie) están construidos de forma tal que la especie perdura, a pesar de que algunos individuos no logren morir de viejos. La idea de adaptación obliga, según creemos, a moverse en estos dos planos: individuo y especie. Esta hipótesis nuclear lleva a una propuesta de una nueva metodología de investigación que Beer denomina "neuroetología computacional":

> El resto de este libro describe un experimento en neuroetología computacional. Un insecto artifcial ha sido desarrollado e insertado en un entorno (simulación) que plantea problemas parecidos a algunos de los que encuentran los insectos naturales. Este insecto artificial es capaz de exhibir algunas de las conductas más básicas requeridas para una supervivencia duradera en el entorno, incluidas locomoción, vagabundeo, retroceso, seguir bordes y alimentarse. Además, es capaz de cambiar apropiadamente entre sus distintas conductas cuando sus condiciones externas e internas cambian. Todo el comportamiento del insecto es producido por un sistema nervioso artificial, cuyo diseño está basado en parte en circuitos neuronales específicos provenientes de distintos animales naturales (pág. 18)

2 Elementos esenciales del modelo Animat

Si se quiere adoptar este modelo en las investigaciones de inteligencia artificial, hay que seguir el siguiente esquema:

1. Establecer de la manera más precisa posible los componentes estructurales que conforman el Entorno. Para lo cual creemos que hay que precisar:

   (1) La estructura de datos propia del Entorno. Por ejemplo, una matriz bidimensional de datos.

   (2) La forma del espacio de datos. Por ejemplo, estableciendo que los límites de dicha estructura serín periódicos. Con ello, si la estructura es unidimensional, la forma del espacio de datos será circular. Toroidal en el caso de una estructura bidimensional.

   (3) Si los elementos de la matriz de datos guardan relaciones de vecindad, especificar el modelo seguido.

   (4) Establecer, caso de que el sistema sea dinámico, cómo se ha establecido el dinamismo. Por ejemplo, si se adopta la perspectiva de un sistema celular, entonces los cambios de configuración del sistema se producirán mediante la computación local de los objetos ubicados en la matriz de datos.

   (5) También hay que declarar si los cambios de configuración se producen sincrónica o diacrónicamente.

2. Definir la estructura del animal artificial. Para ello, creemos que hay que precisar:

   (1) Los objetivos constituyentes o necesidades básicas de dicho animal. Por ejemplo, un objetivo único podría ser que se mueva incesantemente en el entorno, dure lo que dure la simulación.

   (2) Las partes o elementos estructurales que compondrían el animal. Mediante esta estructura, el animal tendría un cuerpo y formaría parte del entorno.

   (3) El esquema básico de dichas partes estructurales es, según creemos, el siguiente:

      1. Receptor o receptores sensitivos
      2. Efector o efectores
      3. Control central o cerebro. Podría, a su vez, estar formado por partes. Aunque no es necesario, sí nos parece conveniente que se usen redes neuronales artificiales.

(4) Aparte debe quedar claro en qué consiste el dinamismo del animal. Por ejemplo, estableciendo un grafo de conexiones entre las distintas partes del animal diseñado.

3 Experimentos en el modelo Animat

Vamos a presentar a continuación el esquema básico de varios experimentos que hemos realizado en el modelo Animat. Seguimos el esquema explicado en la sección anterior: primero presentamos el diseño del entorno y, a continuación, el de animat.

3.1 El entorno

Hemos trabajado con dos diseños básicos: un entorno estático, sin dinamismo; y un entorno dinámico. En ambos casos, en cada elemento de la estructura de datos hay un objeto, incluido animat.

- Estructura de datos: una matriz de datos de f filas y c columnas. En todos los experimentos que estudiaremos, las filas serán 50 y las columnas también 50. En cada posición de la matriz (f_i, c_j) hay un objeto.
- Forma del espacio de datos: la matriz de datos bidimensional tiene límites periódicos. Por esto último, la forma del espacio es toroidal.
- Relaciones de vecindad: Relaciones de vecindad son las de Moore, si bien Animat es estimulado por tres vecinos en cada iteración según su orientación.
- Dinamismo: En el caso de entornos cuyos objetos son a-dinámicos, la ley del dinamismo del Entorno puede expresarse así:

  Si el objeto al que hay que aplicar esta ley no es de la clase animat%, no sucede nada. En otro caso (es decir, si el objeto al que hay que aplicar esta ley es un ejemplar de animat%), cede el dinamismo al objeto de la clase animat%, según las reglas siguientes:

  1. los intercambios de posición con un blanco se asientan inmediatamente en la matriz de datos.
  2. Si la acción de Animat consiste en empujar un objeto, esta acción se realizará si en la posición contigua del objeto empujado hay un blanco. En otro caso, no se realizará el intercambio de posiciones entre el objeto empujado y el blanco contiguo.

- Cambios de configuración: el dinamismo descrito se produce en una secuencia de iteraciones. En cada iteración, el Entorno cambia de *configuración*. En el monitor gráfico se van viendo estas configuraciones sucesivas.[2]

3.2 Animat

En estos experimentos Animat está dotado de receptores sensitivos con los que contacta con el entorno, recibiendo de este tres estímulos. El modelo más básico tiene un único receptor sensitivo. Los estímulos son procesados por una red formada por 15 "fotorreceptores", cada uno de los cuales producirá una respuesta cuyos valores posibles son 0 o 1. La salida de este receptor sensitivo es procesada por una red de tres capas del tipo *perceptrón multicapa*. La salida de esta red es una *orden motora*, que constituye la entrada de los *efectores* de Animat. En el modelo más básico los efectores están constituidos por dos tipos de capacidades: una, de cambio de posición; otra, de giro sobre su propio eje con el consiguiente cambio de orientación. La figura 1 ilustra estas ideas.

Figura 1: Animat: cambio de lugar y giro

3.3 Discusión

3.3.1 El Entorno como totalidad

Cuando se contempla la ventana del monitor, se está viendo el modelo experimental entero, es decir, se está viendo el Entorno como una totalidad entre cuyas partes Animat es una más. Esta totalidad puede ser entendida, en primer lugar, como una colección de objetos.

[2] Para una explicación más extensa remitimos de nuevo al capítulo correspondiente del libro tercero de *Fundamentos de I. A.* de A. Benítez.

Pero esa totalidad, en segundo lugar, tiene una forma determinada. Hemos visto que al Entorno corresponde una matriz bidimensional de datos. Por ello, cada posición queda determinada por un par (fila, columna) que identifica un posible lugar en el espacio del Entorno. El Entorno es así una estructura ordenada de posiciones en la que han de ser dispuestos los objetos. La primera idea de una totalidad como mera colección de objetos ha de ser precisada: el Entorno es una estructura de posiciones en cada de las cuales se ubica un objeto.

Corolario de esto es que dos objetos son distintos, son partes distintas de la totalidad, porque ocupan lugares distintos. Es decir, si suponemos dos objetos que tengan exactamente todos sus propiedades iguales, tendríamos que afirmar que no son el mismo porque han de ocupar dos posiciones distintas. Más aún, si contemplamos un objeto cualquiera en su interior, podríamos afirmar que el estar en una posición no es una propiedad que le pertenezca *per se*, pero en tanto que objeto del Entorno, en cuanto parte de este, cada objeto tiene necesariamente una determinación más —puramente relacional—: ocupar un lugar en el espacio del Entorno.

3.3.2 El dinamismo del Entorno

Hemos visto también que el Entorno, si bien es uno y una totalidad, cambia de *configuración*, esto es, que cambia el *contenido* de la matriz de datos de una iteración a otra. Si se quiere, puede decirse que "puede cambiar". Los cambios que se producen en la configuración del Entorno están *reglados*. Es decir, que existe al menos una regla según la cual se producen los cambios de configuración. Dado que esa regla, o reglas, regula por igual cada cambio, se mantiene constante y uniforme en la secuencia de cambios de configuración. En este sentido, cada regla es "ley".

La regla que rige el dinamismo del entorno puede ser concebida como una única ley: la del intercambio de posiciones. El intercambio puede darse entre dos objetos, un blanco y otro que no puede ser un blanco. En efecto, cuando Animat actúa cambiando de posición se produce un intercambio de posiciones. Cuando Animat empuja un objeto, este se desplaza de lugar y se produce un intercambio de posiciones.

Corolario de lo anterior es que el dinamismo ha de ser concebido como determinación (definición) de las relaciones de posición entre objetos.

3.3.3 Animat como sistema

Animat es un único objeto. Sin embargo, no es un objeto simple, antes bien tiene partes. Es, pues, un objeto compuesto. Como decíamos del Entorno, puede decirse —en primer lugar— que Animat es una colección de partes: receptores, unidad de control y efectores.

En segundo lugar, a diferencia del Entorno, las partes de Animat no tienen una disposición espacial. Sin embargo, sí existe una cierta disposición puesto que en su funcionamiento Animat exige que una parte —receptores— actúe antes de que la que la unidad de control pueda hacerlo. Lo mismo cabe decir de los efectores respecto a la unidad de control. Desde el punto de vista del funcionamiento interno de Animat, las partes tienen una disposición funcional, están ordenadas funcionalmente. Por ello, cada parte que compone Animat es una parte inseparable.

Esta disposición funcional, en tercer lugar, permite determinar cada acción de Animat como una tripla formada por los receptores, la unidad de control y los efectores —siempre en este orden—, según el esquema siguiente:

(estímulos, orden motora, acto efector)

Un ejemplo de acción, desde este punto de vista, podría ser el siguiente:

((blanco, azul, verde), (blanco 0 0 0), avanzar al blanco).

De donde se sigue que cada acción de Animat responde a la idea de *esquema de conducta* de la Etología. Y, según esta idea, es posible determinar las correlaciones entre mecanismos desencadenadores innatos (MDI) y los actos efectores de Animat. En efecto, el acto de avanzar es relativo a la presencia de un blanco entre los estímulos, mientras que el giro se produce cuando no existe un blanco entre los estímulos.

En cuarto lugar, Animat produce una acción en cada iteración. Desde este punto de vista, Animat consiste en actividad incesante. Por ello, Animat no es como el resto de objetos del Entorno, sino que es un objeto cuya existencia entera consiste en producir acciones. Es, pues, un *agente*, si por "agente" se entiende un objeto que produce incesantemente acciones. La "agencia" de Animat queda determinada tanto por las correlaciones posibles entre los MDI y los actos efectores como por las triplas correspondientes a los esquemas de conducta posibles.

Corolario de esto es que Animat es un objeto dinámico, frente al resto de objetos del Entorno que son a-dinámicos.

Animat es, pues, un objeto compuesto; un sistema funcional de partes inseparables; y un sistema dinámico.

REFERENCIAS

Beer, R. (1990). *Intelligence as adaptive behavior: an experiment in computational neuroethology*. Boston: Academic Press.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*, 3–15.

Brooks, R. A. (1991a). Intelligence without reason. Technical Report Memo Num. 1293, MIT.

Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Eibl-Eibesfeldt, I. (1976). *Nueva Antropología. Tomo 2: Antropología biológica*, chapter Adaptaciones filogenéticas en el comportamiento del hombre, (pp. 3–55). Barcelona: Ediciones Omega. Trad. de Margarida Costa.

Eibl-Eibesfeldt, I. (1977). *El hombre preprogramado*. Madrid: Alianza editorial. Trad. esp. Pedro Gálvez.

Eibl-Eibesfeldt, I. (1993). *Biología del comportamiento humano. Manual de etología humana*. Madrid: Alianza editorial. Trad. esp. Francisco Giner y Luis Cencillo.

Wilson, S. W. (1985). Knowledge growth an artificial animal. In Grefenstette, J. (Ed.), *Proceedings of an International Conference on Genetic Algorithms and Their Applications*, (pp. 16–23)., Hillsdale, NJ. Lawrence Erlbaum Assoc.

Wilson, S. W. (1991). The animat path to ai. In Books, T. M. P. (Ed.), *Animals to Animats: Proceedings of The First International Conference on Simulation of Adaptive Behavior*, (pp. 15–21)., Cambridge, MA.

TWO KINDS OF SECOND PERSON AUTHORITY[1]

EDSON BERND (1), ANTONIO GAITÁN (2), MANUEL DE PINEDO (1)
(1) Departamento de Filosofía I, Universidad de Granada
(2) Departamento de Humanidades, Universidad Carlos III

ABSTRACT: Dyadic interactions have become a hot topic of research in both cognitive sciences and philosophy. The perspective we take in such interactions is increasingly regarded as the proper focus to understand basic capacities such as mind reading or evaluative thinking. In this paper we will focus on this perspective from the philosophical side, and we will centre the discussion on four main areas, namely, self-knowledge, theory of mind, reactive attitudes and joint action. We will argue that there is a kind of *constitutive, second-personal,* authority that others have over a subset of our mental states, a kind of authority that is distinct from, and irreducible to, first-person authority.
KEYWORDS: Self-knowledge, first-person authority, second-person authority, reactive emotions, joint intentions

1. The importance of the second-person perspective is clear when we look into two influential debates. The first debate is about self-knowledge, or about the nature of first-personal authority. Here, first-person authority is taken to be the special authority that we concede to the subject over her own thought. This authority is attributed under the assumption that, in normal circumstances, the thinker is the best person to ask about her own beliefs or intentions. Such an assumption is made irrespectively of the mode of access that on any particular occasion one has to her own thought. According to this line of argument, authoritative self-*knowledge*, like all knowledge, rests upon the possibility of comparing perspectives on the same subject matter. Epistemic authority is a property attributed to the person and not to the specific mode of access that she happens to have to her mental life (see Borgoni 2009). Straightforwardly, I, and only I, can have first- *and* third-personal access to my mental states.

This explanation of first person authority uses the contrast between the assumed bipolar structure of first- and third-person perspectives. While third-person perspective stands for the objective view of knowledge, the first-person one stands for the subjective approach. A subjective

[1] Work for this paper was partially funded by the research projects “Disposiciones, holismo y agencia”, “Agencia, normatividad e identidad. La presencia del sujeto en la acción” y “Conceptos de orden superior: Una aproximación expresivista” (Ministerio de Ciencia e Innovación, España, FFI2010-19455, FFI2011-25131 y FFI2010-15704).

approach to knowledge involves something known mainly by means of personal experience, not shared, and normally voiced in a grammatical first-person structure. The subject of knowledge and its mode of access to it are deeply relevant to what is known. An objective view is impersonal, shared and does not depends on the specificities of the particular knower. It is focused on the object to be known independently of the subject of knowledge. The first is based on directness and intimacy, while the second is based on observation.

Is there any way to overcome the above dichotomy between intimacy and observation, between a subjective and an objective stance toward one's own thoughts? Following a Burgean anti-individualist, externalist account of self-knowledge, Heal (2012) introduces a second person perspective as a distinctive route to knowledge of other's mental content. In her account, this perspective can have a constitutive role in the individuation of mental predicates grounded on the social nature of concept acquisition. After arguing that the anti-individualistic view is the best explanation of how we get to know others' mental states, Heal affirms that her theory of co-cognitivism is a natural consequence of adopting an anti-invidualist position. Adopting an anti-individualistic view about mental content leads to adopting a co-cognitive approach to self-knowledge. This approach says that cognition is only possible if shared from the beginning, and the distinctive route suggested by Heal is the participative role of a second person perspective on the individuation of mental content. This role is based on the shared characteristic of the second order cognition of mental states. The point of correction for Heal is that, according to the anti-individualist approach, we normally don't have full grasp of concepts, and meaning remains incomplete in any subject's use. So, when we meet a better knower of a concept we have the opportunity to learn something more about our own usage of this concept.

The other place where the relevance of the second-person perspective is clear is the debate between those who prioritize the objective, third-personal stance, in mental attribution and those who model mindreading by assuming an imaginative, first-personal, stance toward others' mental states. This well-known dichotomy, between Theory-theory and Simulation-theory, dominated for years the discussion inside Theory of Mind. An undesirable result coming from this polarization has been the hiding of a subtle specificity in mental attribution that neither the objective nor the subjective points of view could perspicuously show. The point is that when we interact with beings that we presuppose to be minded we behave differently than when we interact with things or mindless beings. Different authors have aired this dissatisfaction; for instance, Gomila (2002), Pinedo (2004) and Schilbach *et al.* (forthcoming), ranging from a philosophical to a neuroscientific outlook. They propose a fundamental role for the second person point of view in the construal of

what social cognition is. Peter Strawson (1974), as we will show below, can be seen as a precursor of this view in his argument for the essentially interpersonal character of reactive emotions (even though he uses them in the context of a discussion unrelated to that of self-knowledge).

2. Following these debates, the introduction of a second person perspective in the contemporary discussion recommends an understanding of interaction akin to shared attention: being "jointly-attentive" is to be attentive to an object that other subject is also attending to and, moreover, being attentive to her attention to it, and vice versa. This is roughly Naomi Eilan's (2005) conception in her influential paper "Joint attention": "[...] everything about the fact that both subjects are attending to the same object is out in the open, manifest to both participants." (p.1) This bipolar structure can also be found in Donald Davidson's famous appeal to triangulation while interpreting others' linguistic performances (Davidson 2001), Michael Tomasello's findings about 'a special ability to identify with others' (Tomasello 1999. p. 77), Daniel Hutto's narrative account of folk-psychology and, more recently, Laurent Schilbach's claim that second personal engagements are necessary for language acquisition (Schilbach *et al.*, in press).

All these accounts stress the methodological importance of a second-person perspective. Our topic, however, will be second-person *authority*. For instance, sometimes an old friend may know better than us what we really like, intend, prefer, etc. From this privileged standpoint, she can help us recover *self*-knowledge, showing us that we are wrong about the real state of mind we presumed we were in; and show us that we are really in another state. Keeping in mind these cases, we will distinguish in the rest of this paper two kinds of authority that we can attribute to anyone occupying an analogous role to our old friend in the above example.

There is a *weak sense* of second-person authority according to which some people have authority over the *content* of some of our thoughts. This kind of second-person authority is a consequence of the special familiarity that the second person has with respect to my thought. It is easy to find situations where the second person is in a position to know better than myself the content of my beliefs, intentions or feelings not in virtue of displaying a detached, observational and inferential access to my thinking, but rather in virtue of her intimate knowledge of it. This authority potentially extends to the contents of all types of mental states: a person that knows me well can, on occasions, be in a better position than myself to know what I belief, want, intend or feel.

And there is a *strong sense* of second-person authority according to which second-personal engagements are essential for the constitution (both ontogenetic and phylogenetic) of certain *types* of mental *attitudes*. We will now focus on two kinds of attitudes: intentions linked to joint action

and reactive emotions. Reactive emotions –such as resentment or gratitude– were first highlighted by Peter Strawson within the context of the debate between freedom and determinism. Strawson holds that these reactive emotions are only intelligible in a framework where affection, esteem, contempt or malevolence towards oneself are recognizable by whoever has such emotions. The same goes for joint intentions. For a joint intention to succeed, it is necessary that *we both recognize and reinforce the other's intention* to engage in a shared enterprise. It is essential to such shared intentions that they are knowable by someone else.

Our proposal concerning the existence of a constitutive or strong second-person authority is not subject to a possible objection that could be raised against a different conception of such an authority: the second person is not a mere representative of the community in the process of introducing the individual into the practices of the group; someone better placed than the rest, so to speak, to have a privileged, non-inferential access to the content of some of our mental states. This "weak" authority (the one that Heal attributes to the second person) is grounded both on (a) the collective rules of application of the words to express a more or less specific mental state; (b) the special acquaintance the second person has with us that gives her a privileged cognitive position compared to other competent users of the same language. Her authority is constituted by these two elements. We call it "weak" second-person authority because it only reaches the *content* of our mental states, rather than helping constitute some types of mental *attitudes*. In the next and final section we further explore two instances of this "strong", constitutive second-person authority.

3. As we have said, strong second-person authority is linked, at the very least, with reactive emotions and intentions behind joint action. Strawson pointed out that we have two different basic attitudes in social relationships: an "objective attitude" and a "participant reactive attitude". The objective attitude sees the others as "objects" of social interaction, as things we need to understand for prediction and control. The "participant reactive attitude" considers the others as persons, as individuals of the same type as the subject. For Strawson, we normally start with participant attitudes, expecting others to meet a basic demand of good will or consideration. When this basic demand isn't satisfied we feel resentment. The important aspect of this event, the action that triggers the reactive emotion, is that the subject was expecting an appropriate behavior from the offender in particular. Our reactive attitude is controlled by our view of her specific offending action or even by our view about her as an agent.

Resentment depends on a complicated script of second-personal mind attribution (see Darwall 2006 for a similar argument linking reactive emotion and second-person perspective).

When someone teaches us how to behave and which behavior we should expect from others, we also learn how to feel about their actions and how to evaluate whether it is in accordance with the script of possible appropriate actions or not. So, it is an emotion that depends on interpreting her act as being directed to us, personally, directly, towards us, and only us, in that specific event. This is a second-personal engagement. The person who teaches us the concept of resentment teaches us from a position of second-person authority. The strength of this sense of second-person authority is that someone else constitutes a mental state in us. Our "teacher" creates it from a basic state of frustration into a complicate feeling of resentment, an emotion we can only feel by learning how to interact with others. Our mother or our caregiver is the person who is in a position of authority over our mental life in a very strong sense, a sense in which the second person is responsible for bringing us up well.

Another example of a strong sense of second person-authority can be found between agents sharing an intention to act together (Sebastian Rödl 2010). People are jointly acting when: (a) both of them want the goal to be reached; (b) each one of them do different parts of the set of particular partial actions that leads to the conclusion of the act, reaching the goal; (c) each one knows that by acting together they are together reaching the goal; (d) each one knows what part the other is doing and knows which is her own role in the action, and (e) each one knows that the other knows, and is expecting the other to fulfill its part while committing oneself to do one's own part. People are acting together if and only if they know they are acting together.

In Rödl's example, Mark and Helen are making tea together. Helen is filling the kettle, while Mark is lighting the stove. Both actions are part of an act, namely, to make tea, and could be performed by the same person. What give its significance and its cause is the intention to bring about the goal, namely the tea. If Helen was filling the kettle and were asked "what are you doing?", her answer would not be "I am filling the kettle!". This reply sounds comical because it's obvious and doesn't answer the question. When someone sees someone else performing an action and asks what is she doing, the question is about the goal: What are you doing (in filling the kettle, for instance)? The right answer then would be "I'm making tea". The same answer would be appropriate if the question was about lighting the stove. Then, what defines the action is its goal. If these two actions were performed by two people, like Mark and Helen, the answer to what were they doing would be "making tea". So, there is an action whose parts are, among others, filling the kettle and lighting the stove, all of them defined by its goal, which is making tea. Nevertheless, to be a joint action, this set of partial actions, combined as an act, needs to be done by the two agents knowing that they are doing the partial acts as combined to reach the goal. Moreover, they need to

know that the other is doing her/his part of the act by performing the partial action. Rödl makes use of Anscombe's thesis that someone is doing something intentionally only if she knows that she is. So, for an action to be classified as a joint action it must be conceived as an action aiming at a goal, known by all the agents involved, as the intention of the general action, accomplished by means of the partial acts performed by each one of the agents. The agents know that all of them are performing specific movements to reach a common goal at the end.

Now, we can focus on the particular second-person authority revealed by the possible mutual demand on each other to execute their particular actions that lead to the achievement of their common goal. Helen and Mark have mutual expectations on each other. Helen expects from Mark that, while she is filling the kettle, he is lighting the stove, and reversely, Mark expects form Helen that while he's doing so she is filling the kettle. Normally, in a situation like this, both of them are confident that the other is doing its part. However, if Mark stops his action, for whatever reason, Helen is entitled to require from him that he does his part. This demand and its entitlement are grounded on a kind of second-person authority, similar to the moral second-person authority proposed by Darwall (2006). The main difference between our proposal and Darwall's is that in his proposal the authority is supported by generic moral assumptions, while in ours the authority rests on the particular, even if tacit, agreement between the specific agents committed to reach together the goal to be achieved.

We maintain that this is a kind of strong second-person authority because it is constituted by the specific relation between two individuals that entitles each other to demand specific actions to be performed. It would make no sense for anybody else who was not part of the agreement to make demands from any of them. Nobody else is qualified to require it. There is no third-personal general reference to collective underlying rules as sources of normativity, beyond, of course, the rules of language use, which have no direct influence either in the agreement itself or in its execution. The entitlement of both participant members of the agreement rests solely on the achievement of the common goal. In this case, unlike the above case of strong second-person authority, there is no asymmetry between requesting and requested agent. Both of them are in a privileged position to command a specific action on the other.

REFERENCES

Borgoni, C. 2009, *Knowing the Nature of One's Mind: An Externalist Basis for Self-Knowledge,* Tesis doctoral, Universidad de Granada.
Darwall, S. 2006, *The Second Person Standpoint*, Cambridge, Mass., Harvard University Press.
Davidson, D. 2001, *Subjective, Intersubjective, Objective*, Oxford, Clarendon Press.

Eilan, N. 2005, "Joint attention, communication and mind", N. Eilan *et al.* (eds.) *Joint Attention: Communication and Other Minds, Problems in Philosophy and Psychology,* Oxford, Oxford University Press.

Gomila, A. 2002, " La perspectiva de segunda persona de la atribución mental", *Azafea, Revista de Filosofía,* 4, 123-38.

Heal, J. 2012, "Social anti-individualism, co-cognitivism and second person authority", *http://cas.uchicago.edu/workshops/wittgenstein/files/2011/09/Social-anti-individualism.pdf*

Hutto, D. 2008, *Folk Psychological Narratives. The Socio-cultural Basis of Understanding Reasons*, Cambridge, Mass., MIT Press.

Pinedo, M., "De la interpretación radical a la fusión de horizontes: La perspectiva de la segunda persona", in J. J. Acero *et al.* (eds.), *El legado de Gadamer*, Granada, Editorial de la Universidad de Granada, 225-35.

Rödl, S. 2011, "Joint action and plural self-consciousness" in *http://cas.uchicago.edu/workshops/germanphilosophy/2011/10/08/oct-13-14-sebastian-rodl/*.

Schilbach, L. *et al.* (in press), "Toward a second-person neuroscience", *Behavioral and Brain Sciences*.

Strawson, P. 1974, "Freedom and Resentment", in *Freedom and Resentment and Other Essays*, London, Routledge, 2008, 1-28.

Tomasello, M. 1999, *The Cultural Origins of Human Cognition*, Cambridge, Harvard University Press.

¿FUE WHORF UN NEOWHORFIANO?[1]

ANTONIO BLANCO SALGUEIRO
Universidad Complutense de Madrid
ablancos@filos.ucm.es

ABSTRACT: The aim of this paper is to suggest a weak interpretation of Whorf's ideas, in line with recent "neowhorfian" approaches. Various Whorf-inspired hypotheses about the impact of language on though are examined and some of them are found both plausible and interesting. In the final section, some connections between the hypotheses are explored.

KEY WORDS: Whorf, neowhorfianism, language, thought

Introducción

Tras un período de ortodoxia universalista, asistimos a una revitalización del whorfismo. Los "neowhorfianos" suelen apostar por versiones débiles de la tesis de que el lenguaje (o sus concreciones, las lenguas) tiene un impacto sobre el pensamiento y tienden a ofrecer lecturas también moderadas de Whorf.[2] La obra de Whorf es compleja, fragmentaria y sólo parcialmente publicada, lo que hace difícil extraer una formulación consistente que refleje lo que "realmente" dijo o quiso decir. Ello ha permitido a sus detractores reconstruir sus ideas de modos que las hacen parecer tan radicales que son fácilmente refutables. La moderación neowhorfiana promueve también interpretaciones en el otro extremo: los efectos de las lenguas sobre el pensamiento y la conducta son triviales, luego el defensor de la autonomía y universalidad del pensamiento no tiene de qué preocuparse. La viabilidad del neowhorfismo depende pues de formular hipótesis acerca del impacto del lenguaje sobre el pensamiento que sean a la vez *plausibles* (refutables, en todo caso, aportando evidencia detallada en contra) e *interesantes* (que apunten a efectos sorprendentes, que no se sigan de la propia función comunicativa del lenguaje). Las formas de moderación que evitan la trivialidad beben de tres fuentes principales: 1) La posible débil *intensidad* y fácil *reversibilidad* de los efectos, que ilustrarían la *influencia* de la lengua sobre el pensamiento, frente a las lecturas *deterministas* de Whorf. 2) El *alcance* del impacto cognitivo de las lenguas, que puede limitarse a

[1] Esta investigación se realizó en el marco del proyecto FFI2009-08828/FISO, financiado por el MICINN.

[2] Tres interpretaciones recientes de Whorf son Schultz (1990), Lucy (1992) y Lee (1996). Algunas buenas introducciones a los debates actuales sobre la relación pensamiento-lenguaje son Gumperz & Levinson (eds.) (1996), Carruthers & Boucher (eds.) (1998), Pütz & Verspoor (eds.) (2000), Niemeier & Dirven (eds.) (2000), Gentner & Goldin-Meadow (eds.) (2003) y Malt & Wolff (eds.) (2010).

ciertos dominios mentales, frente a la tesis de un efecto *global*. 3) La posible *universalidad* de los efectos, que lo serían así del *lenguaje,* más que de las distintas lenguas. Aclarar esas formas de moderación permite reconstruir el espacio lógico de posibles respuestas al problema de la relación lenguaje-pensamiento.

Trataré de extraer de la obra de Whorf distintas tesis moderadas e interesantes. La reflexión abarcará siete puntos: 1) Las vacilaciones en la denominación de la relación lenguaje-pensamiento. 2) La importancia de la noción de hábito. 3) El carácter atribuido a la experiencia prelingüística. 4) El papel de las metáforas lingüísticas en el pensamiento. 5) La concepción holística del lenguaje y del pensamiento, así como de su relación. 6) La dirección de la relación entre lengua y pensamiento (y cultura): unilateral *vs.* bilateral. 7) La influencia del habla en el pensamiento.

2. Problemas con la terminología básica

Algo que exaspera en Whorf (1939b) es la cantidad de términos no sinónimos usados para referirse a la relación de una lengua con el pensamiento y la conducta de sus hablantes. El más recurrente es "influencia", que muchos neowhorfianos usan para marcar su moderación, frente al *determinismo* lingüístico a menudo atribuido a Whorf. Otras veces emplea el neutral "afecta". Pero usa también expresiones más radicales como "condicionamiento", "compulsión", "determinación", "dependencia", etc., aunque a menudo acompañadas de cualificaciones mitigadoras, como al afirmar que el pensamiento está condicionado *en parte* por la lengua nativa, lo cual o es de nuevo algo más débil (una influencia), o implica que hay otros factores que también condicionan *en parte* (la parte restante) el pensamiento, o apunta a una visión dialéctica en la cual el pensamiento en parte afecta y en parte es afectado por el lenguaje.

3. El pensamiento habitual

Un motivo central en Whorf (1939b) es que la lengua (su uso constante) afecta al pensamiento y a la conducta *habituales*. Esa posición, sin ser trivial, no resulta radical.[3] Detrás de la noción de hábito late la de *reversibilidad* de los posibles efectos, que se opondría a la idea de su carácter permanente u obligatorio. Eso vale para la "visión del mundo" supuestamente promocionada por la lengua materna. Whorf habla de "mundos habituales de pensamiento", lo que abre la posibilidad de que podamos introducirnos en mundos no habituales, y explicaría que el lingüista pueda penetrar en el mundo "abierto" por una lengua extraña, o que los científicos sean

[3] Y podría resultar falsa. Munich & Landau (2003) rechazan efectos de las diferencias lingüísticas en los modos habituales de pensar en una variedad de dominios en los que los neowhorfianos los postulan.

capaces de adquirir nuevos modos de hablar que se ajusten a visiones novedosas del universo; en general, protegería a Whorf contra preguntas a las que difícilmente pueden responder las versiones radicales del relativismo lingüístico: ¿cómo es posible el *cambio lingüístico* si el pensamiento está encerrado en una lengua particular?, ¿cómo se establecen, *en primer lugar*, los esquemas lingüísticos que afectan al pensamiento, esto es, cómo se originan las lenguas (en particular, las diferencias entre ellas)? La respuesta podría ser: con esfuerzo reflexivo, el que se necesita para crear ideas que se salgan de los caminos trillados y que con el tiempo pueden volverse corrientes; o para responder a nuevas circunstancias que desafían o hacen inútiles los viejos hábitos y las viejas palabras. Esa interpretación "influencista" queda reforzada por tres consideraciones: 1) La insistencia en que la lengua afecta, sobre todo, al pensamiento inconsciente (funciona al nivel del trasfondo). 2) La afirmación de que el lenguaje "representa la mente de la masa", la más conservadora, la menos consciente de sus sesgos y la menos dispuesta a admitir que hay formas alternativas de hablar-pensar. 3) La alusión a los beneficios cognitivos de la consciencia metalingüística, que permitiría controlar y modificar los hábitos lingüístico-cognitivos. Esa versión de la tesis relativista puede formularse así:

TRL_1 *La lengua fomenta en sus hablantes nativos hábitos cognitivos y conductuales no universales (distintivos de cada lengua particular).*

4. La experiencia pre-lingüística no es un flujo caleidoscópico

Existe una diferencia entre Whorf (1939b) y Whorf (1940) en cuanto a la complejidad atribuida a la experiencia prelingüística universal a la especie, sobre la que actuarían las lenguas para generar imágenes del mundo alternativas. Un célebre pasaje de Whorf (1940) afirma que sin el lenguaje la experiencia es un "flujo caleidoscópico", lo que sugiere su carácter amorfo. En cambio, Whorf (1939b) alude constantemente al carácter diferenciado de esa experiencia. Así, por ejemplo, afirma que hay un sentido básico y prelingüístico del tiempo ("la esencia del tiempo"), que es la experiencia del hacerse más y más tarde. Además, reivindica la visión hopi del mundo como más adecuada que la europea precisamente por su supuesta mayor fidelidad a la experiencia prelingüística. Tenemos entonces aquí una versión matizada (plausible y no trivial) de la relatividad lingüística, compatible con formas moderadas de innatismo conceptual:

TRL_2 *La lengua contribuye (en mayor o menor medida) a alterar, elaborar o reconfigurar la estructura pre-lingüística innata de la experiencia humana.*

La importancia de este nivel universal de representación (ligado al interés de Whorf en la psicología de la Gestalt) es grande, ya que en él funda la posibilidad de comparar lenguas sin presuponer el marco conceptual de una de ellas (la del propio lingüista) (Whorf 1939a; Whorf & Trager 1938). Detrás de los dibujos con los que Whorf acompaña a veces sus contrastes entre lenguas está el presupuesto de un marco experiencial neutral, la pervivencia en los hablantes de una capacidad para el pensamiento figurativo ("gestáltico") transversal a las lenguas que posibilitaría la penetración en los modos alternativos de describir una situación.

Por supuesto, la visión whorfiana sobre qué conceptos se ven afectados por la lengua podría estar errada. Así, afirma que conceptos como 'tiempo', 'espacio', 'causa', 'efecto', 'progreso', 'el pasado', 'el futuro', 'sustancia', 'materia' (que siguen siendo los más explorados por los neowhorfianos) no son universales, sino específicos de la "metafísica implícita" de las lenguas europeas (Whorf & Trager 1938). Pero, por otra parte, llevado por su misticismo, considera universales partes de la mentalidad humana como el pensamiento religioso, ético y estético (*ibid.*, 272). Lo importante es la hipótesis de que algunos conceptos básicos varían debido a diferencias lingüísticas, algo muy distinto de la afirmación a menudo asociada con Whorf de que la imagen del mundo es, en todos sus detalles y sin restricciones debidas a la realidad objetiva o a la naturaleza humana, una construcción lingüística.[4]

5. El poder de las metáforas

Según Whorf (1939b), la influencia del lenguaje se produce, en buena medida, a través de ciertas metáforas convencionales, en particular *metáforas espaciales*, de un modo que recuerda (e inspira) las ideas de Johnson & Lakoff (1980), dejando de lado que Whorf enfatiza el poder distorsionador de las metáforas, mientras que Johnson & Lakoff consideran que gracias al pensamiento metafórico podemos conceptualizar dominios abstractos (alejados de nuestra experiencia básica). Según Whorf, el tiempo occidental es pensado habitualmente según metáforas espaciales ausentes en hopi. Si este fuera el mecanismo principal mediante el cual el lenguaje afecta al pensamiento tendríamos de nuevo una tesis débil, ya que no parece difícil abandonar las viejas metáforas (al menos no tanto como cambiar aspectos básicos de la gramática), o dejarse guiar por metáforas novedosas. Johnson & Lakoff señalan que muchas lenguas no sólo conceptualizan el tiempo de acuerdo con metáforas espaciales,[5] sino de acuerdo con metáforas convencionales como

[4] Boroditsky (2001) sugiere, en línea whorfiana, que es más probable que la lengua influya en dominios mentales abstractos como el tiempo que en dominios cercanos a la experiencia sensorial como el color.

[5] Dentro de las metáforas espaciales, Boroditsky (2001) señala que el Mandarín emplea una metáfora vertical, distinta de la metáfora horizontal central en lenguas como el inglés (o el español).

"el tiempo es dinero", por lo que algo puede costarnos tiempo, y por lo que nuestro tiempo se puede perder, gastar, ahorrar, invertir, prestar, etc. Esta nueva tesis de inspiración whorfiana podría expresarse así:

TRL$_3$ *Las metáforas que se utilizan (de un modo habitual y sin consciencia de su carácter metafórico) para hablar de ciertos dominios varían de lengua a lengua y, con ellas, los modos de pensar en esos mismos dominios.*

6. Holismo

Cuando Whorf contrasta el tratamiento del tiempo en las lenguas europeas con el tratamiento hopi, uno pensaría que va a comparar los respectivos sistemas verbales. Sin embargo, empieza encontrando la diferencia en el modo en que usan el plural. Según él, los hopi no usan el plural para hablar del tiempo o de fenómenos cíclicos. En general, defiende que es la lengua como un todo integrado, y no aspectos parciales de la misma, la responsable de los efectos cognitivos. La noción de *fashion of speaking* sirve para resaltar el carácter holístico de los efectos de la lengua sobre el pensamiento. También defiende la tesis complementaria de que, dado que el pensamiento es un todo integrado, la influencia de la lengua en un dominio mental se extiende a otros. La tesis sería:

TRL$_4$ La lengua como un todo influye en el pensamiento como un todo.

Una derivación de este holismo puede conducir a posturas poco plausibles. Lakoff (1987) reprocha a Whorf que presente las lenguas como todos monolíticos que permiten una única conceptualización de la realidad. Las lenguas son flexibles, permiten múltiples perspectivas sobre los fenómenos, que pueden ser contemplados de modos alternativos sin cambiar de lengua. En concreto, contienen diversas metáforas que resaltan u ocultan aspectos diferentes. Así, la idea de una conversación *larga* (metáfora espacial) y la de *malgastar* el tiempo (metáfora del tiempo como dinero o como recurso valioso) pueden aplicarse a una misma situación. El reconocimiento de la adaptabilidad de los sistemas conceptuales implícitos en una lengua a las intenciones de los hablantes permite desembarazarse de la idea popular (sugerida por el holismo whorfiano) de que aceptar la influencia de las lenguas sobre el pensamiento implica aceptar que los hablantes monolingües están encerrados en sus respectivas cárceles lingüísticas (y tal vez podría hacer compatible el whorfismo con el intencionalismo griceano).

7. Efectos unilaterales vs. influencia dialéctica

Whorf (1939b) debilita la intensidad de los supuestos efectos de la lengua sobre la cultura, con respecto a su poder sobre los individuos. En ambos casos (lengua-cultura y lengua-pensamiento) admite efectos bilaterales. Pero la postura final no es la de una dialéctica entre iguales, sino la de una *dialéctica asimétrica*: factores que interactúan entre sí, pero en la cual uno de ellos (la lengua) tiene más fuerza que el otro:

TRL$_5$ *La lengua no sólo afecta al pensamiento, la conducta y la cultura, sino que es a su vez afectada por ellas, pero la influencia de la lengua sobre esos otros factores es mayor que la que ellos ejercen sobre la lengua.*

8. ¿Efectos de la lengua o del habla?

Las páginas más célebres de Whorf (1939b) son las primeras, en las que presenta evidencia tomada de su experiencia como técnico en prevención de incendios y afirma que los modos de describir ciertas situaciones indujeron las acciones que provocaron los accidentes. Aquí parece apuntarse a una tesis que tiene poco que ver con TRL, entendida como la tesis de que la *lengua* (esto es, su asimilación por parte de los hablantes) afecta al pensamiento. Algunas ideas populares sobre la influencia del lenguaje sobre el pensamiento, como las que subrayan el uso eufemístico de las palabras, apuntan a efectos del *habla* (del *uso* de una lengua). No existe nada en la lengua misma que impida describir una acción militar como “crimen de guerra” y no como “daños colaterales”. La tesis resultante sería:

TRL$_6$ El modo en que hablamos (o nos hablan) de una situación afecta a (influye en, pero no determina) cómo la pensamos.

Por otra parte, los efectos del habla podrían no ser banales si se conectan con lo dicho acerca de los hábitos cognitivos generados por los hábitos lingüísticos. Por mucho que la lengua posea multitud de formas para hablar de ciertos tipos de situación, si es usual hablar de ellas de cierto modo y se descartan, en la práctica, otros modos posibles, ello podría acarrear inercias cognitivas similares a las que se producirían si la lengua misma careciera de recursos expresivos para describir la situación de modos alternativos.

9. Consideraciones finales

Con independencia de la postura real de Whorf, su obra sugiere distintas tesis moderadas pero nada triviales que deben ser confirmadas o refutadas empíricamente. Por supuesto, no he pretendido agotar las tesis de inspiración whorfiana dignas de ser exploradas, ni sugerir que debamos limitarnos a explorar tesis prefiguradas por Whorf.

En cuanto a cómo se relacionan las tesis (y otras que pudieran añadírseles), podemos señalar lo siguiente: 1) Es engañoso hablar de *la* relación entre lenguaje y pensamiento, ya que es probable que el vínculo entre ellos sea complejo y que existan distintas relaciones relevantes. 2) Una vez que se señalan varias dimensiones, se puede investigar cómo se relacionan entre ellas. Por ejemplo, la idea de que parte de la influencia del lenguaje sobre el pensamiento es a través de metáforas convencionales puede contribuir a explicar por qué esa influencia tiene la forma de hábitos reversibles: las metáforas no parecen difíciles de abandonar o de extender creativamente (algo similar sucedería si los efectos se produjesen principalmente al nivel del habla). También podría defenderse que el que las metáforas utilicen conceptos extraídos de la experiencia universal y básica para conceptualizar dominios abstractos explica por qué encontramos relatividad precisamente en esos dominios. Por último, la presión de la experiencia prelingüística y la capacidad metacognitiva de contrarrestar los efectos (en forma de hábitos cognitivos) del lenguaje se complementan a la hora de establecer límites (podríamos denominarlos "inferior" y "superior") a la influencia del lenguaje sobre el pensamiento.

Teniendo en cuenta la complejidad de la cuestión de la relación lenguaje-pensamiento, la filosofía debería contribuir, al menos, a la aclaración conceptual del espacio lógico de posibles respuestas. Sin embargo, muchos filósofos tienden a situarse en las posiciones universalistas o relativistas más extremas, y pierden el interés cuando se les señala que existe una infinidad de posturas matizadas. Por ejemplo, que puede que sólo una parte del pensamiento sea afectada por el lenguaje, o que los efectos serán tal vez leves y reversibles, o que las partes afectadas lo pueden ser de distintos modos, o en distintos grados, o que unos efectos puedan ser cuantitativos –aumento de la atención, o de la memoria- y otros cualitativos, etc. Aquí se les podría aplicar lo que dijo Austin en un contexto diferente: "no son las cosas, son los filósofos los que son simples".

REFERENCIAS

Boroditsky, L. 2001, 'Does language shape thought?: Mandarin and English speaker's conceptions of time', *Cognitive Psychology* 43/1, 1-22.

Gentner, D. y S. Goldin-Meadow 2003, *Language in Mind. Advances in the Study of Language and Thought*, Cambridge, Mass., MIT Press.

Gumperz, J. J. y S. C. Levinson 1996, *Rethinking Linguistic Relativity,* Cambridge, Cambridge University Press.

Lee, P. 1996, *The Whorf Theory Complex: A Critical Reconstruction*, Amsterdam-Philadelphia, John Benjamins.
Lakoff, G. 1987, *Women, Fire and Dangerous things,* Chicago, University of Chicago Press.
Lakoff, G. y M. Johnson 1986, *Metaphors We Live By*, Chicago, University of Chicago Press.
Lucy, J. A. 1992, *Language Diversity and Thought: A Reformulation of the Linguistic Relativity Hypothesis,* Nueva York, Cambridge University Press.
Malt, B. C. y P. Wolff 2010, *Words and the Mind,* Oxford, Oxford University Press.
Munnich, E. y B. Landau 2003, 'The effects of spatial language on spatial representation: setting some boundaries', en Gentner & Goldin-Meadow, 113-155.
Niemeier, S. y R. Dirven (eds.) (2000): *Evidence for Linguistic Relativity,* Amsterdam, John Benjamins.
Pütz, M. y M. Verspoor (eds.) 2000, *Explorations in Linguistic Relativity,* Amsterdam, John Benjamins.
Schultz, E. A. 1990, *Dialogue at the Margins. Whorf, Bakhtin, and Linguistic Relativity,* Madison, University of Wisconsin Press.
Whorf, B. L. 1939a, 'Gestalt technique of stem composition in Shawnee', en Whorf (1956), 160-172.
- 1939b, 'The relation of habitual thought and behaviour to language', en Whorf (1956), 134-159.
- 1940, 'Science and Linguistics', en Whorf (1956), 207-219.
- 1956, *Language, Thought and Reality*, Cambridge, Mass., MIT Press.
Whorf, B. L. y G. L. Trager (1938): 'The 'Yale report'', apéndice a Lee (1996), 251-280.

LO QUE SE DICE Y "DECIR LO MISMO"

DAVID BORDONABA and NEFTALÍ VILLANUEVA
UNiversidad de Granada

ABSTRACT: the aim of this paper is to show some shortcomings on certain criteria to individuate *what is said*. Our strategy is twofold: First, we will analyze the criterion of *what is said* proposed by I Stojanovic in her latest book, *What is said* (Stojanovic 2009), and show how it fits within the theoretical landscape on the matter. Secondly, some counterexamples, will be provided in order to show that the aforementioned criterion is both too strict and too liberal.
KEYWORDS: *what is said*, minimalism, contextualism, Stojanovic.

Introducción

El objetivo de este trabajo es discutir uno de los criterios propuestos por Isidora Stojanovic en su libro *What is said* (Stojanovic 2008, 105-121) para individuar *lo que se dice* (la proposición expresada) y colocar su propuesta en el marco del debate general que tiene por objeto la individuación de los distintos niveles proposicionales de contenido asociados con nuestras proferencias comunicativas.

1. Deícticos y proposición expresada

La estrategia en cuestión, a la que a partir de ahora llamaremos "Decir lo mismo", es la siguiente: si podemos decir con verdad que dos proferencias dicen lo mismo, los elementos que formarán parte de *lo que se dice* serán aquellos elementos comunes a ambas proferencias; *lo que se dice* será neutral con respecto al resto de elementos –lingüísticamente determinados o puramente contexto-dependientes– que contengan las oraciones en cuestión consideradas en su contexto de proferencia. La citada autora utiliza esta estrategia para defender en particular que las referencias de las expresiones deícticas no forman parte de *lo que se dice*, aunque las consecuencias de la misma, como veremos, tienen mayor alcance. Nuestro objetivo será mostrar que dicha estrategia no es adecuada como criterio para individuar *lo que se dice*, ya que puede aplicarse a casos donde los resultados generados son claramente contra-intuitivos. Por un lado, es demasiado *liberal*, dado que admite dentro de *lo que se dice* constituyentes que habitualmente uno querría situar como parte de otros estratos de contenido proposicional asociados con nuestras proferencias, y, por otro, demasiado *estricta*, puesto que deja fuera de *lo que se dice* elementos habitualmente asociados con

el nivel de significado proposicional que nos interesa y cuya exclusión es difícilmente justificable.

El objeto de la obra citada es mostrar que las expresiones deícticas no contribuyen a *lo que se dice* con sus referencias, sino con algo más básico, su significado léxico. Una de las estrategias para mostrar esto consiste en defender que solo forma parte de *lo que se dice* el material común entre los significados de dos proferencias de las que podamos decir con verdad que "dicen lo mismo". Esta estrategia puede ilustrarse a través del siguiente ejemplo:

(1) Estoy escribiendo un libro sobre Montague (dicho por el Prof. Feferman).
(2) Estoy escribiendo un libro sobre Montague (dicho por Inma).
(3) Eso es lo que dijo el Prof. Feferman (posible réplica a (2)).

La proferencia de (3) en un contexto apropiado puede generar una ambigüedad entre una interpretación *estricta* (el Prof. Feferman dijo que *Inma* estaba escribiendo un libro) y una interpretación *flexible* (el Prof. Feferman dijo que *él mismo* estaba escribiendo un libro). Si en el contexto de proferencia de (3) se estableciera como conocimiento compartido entre hablante y oyente que el Prof. Feferman no conoce a Inma, la única interpretación disponible de (3) sería la interpretación *flexible*. En este caso, las referencias de los pronombres de primera persona elididos en (1) y (2) son distintas, pero aún así tenemos la intuición de que es correcto afirmar que en ambos casos se ha dicho lo mismo. Así, no parece que sea necesario incluir dichas referencias en *lo que se dice* para poder afirmar que se ha *dicho* lo mismo. De esto se puede inferir, como hace la autora, que *lo que se dice* es neutral con respecto a las referencias de las expresiones deícticas contenidas en (1) y (2); que podemos afirmar con verdad (3) a pesar de que los pronombres elididos en ambas oraciones refieran a individuos diferentes, las contribuciones de las expresiones deícticas al significado de nuestra proferencia no alteran *lo que se dice*.

Esta estrategia es solo un mecanismo, aunque uno realmente importante, que da apoyo a una de las ideas principales de la citada obra: el compromiso con que *lo que se dice* contiene exclusivamente la cantidad mínima de contenido común entre ambas interpretaciones.

2. ¿Por qué *lo que se dice*?

Antes de presentar nuestros argumentos contra el criterio propuesto por I Stojanovic en la mencionada obra, a lo largo de las dos secciones siguientes, presentaremos dos ideas preliminares: i) la importancia de la noción de *lo que se dice* en la filosofía del lenguaje actual y ii) el marco en el que se circunscribe la posición que se defiende en la obra.

El resultado de lo que hacemos cuando hablamos, el resultado de nuestras proferencias

comunicativas verbales, puede clasificarse de muchas maneras. Puede interesarnos, por ejemplo, durante cuánto tiempo estamos profiriendo sonidos, la altura, el timbre o la intensidad de los mismos, etc. En muchas ocasiones seguramente lo más razonable es prestar atención a estos rasgos del discurso y olvidarse de lo demás. En otras, sin embargo, estamos interesados en determinar la información asociada con los actos comunicativos. Esta información es de muy diversa índole, pero una de las categorías que suelen centrar nuestra atención dentro de este conjunto de información es la de la información de carácter *proposicional*, la que es susceptible de ser declarada verdadera y falsa. Esta última no forma a su vez un todo monolítico. Hay un conjunto potencialmente ilimitado de proposiciones asociadas con nuestros actos comunicativos. En algunos casos somos conscientes de la posibilidad de que se deriven estas proposiciones, en otros no. Podemos suponer que un hablante competente del castellano, por ejemplo, nos pasará la sal cuando proferimos 'pásame la sal' en un contexto apropiado. A su vez, podemos suponer que el oyente asumirá que hay sal en algún lugar próximo, que hay algo que le estamos pidiendo, que estamos comiendo, que nuestra comida necesita sal, etc. Podemos incluso, si el contexto es propicio, anticipar que el oyente puede ofenderse al sentirse tratado como un sirviente al que se le pide que lleve a cabo una tarea perfectamente al alcance del que habla. Podríamos incluso aventurar que puede haber información que nuestra proferencia sugiera al oyente que está fuera de nuestro control a pesar de poseer también carácter proposicional. Al oír la palabra sal, por ejemplo, un oyente informado puede pensar en algunos versos de Alberti, un gaditano podría quizás pensar en la gracia característica de la gente de su tierra –el salero–, y a un melómano puede venírsele a la cabeza una declaración cantada de Camarón de la Isla.

Obviamente, con respecto a todo este contenido de carácter proposicional, de un modo u otro asociado con nuestra proferencia, no exhibimos el mismo compromiso. La facilidad con la que podemos, por ejemplo, retractarnos de estas proposiciones cambia enormemente. Con respecto a muchas de ellas ni siquiera tiene sentido pensar en la posibilidad de echarnos atrás, porque no están bajo nuestro control. Podemos arrepentirnos de haber pedido la sal en lugar del azúcar, explicar que en realidad no estamos comiendo y queremos la sal para cualquier otra cosa, aclarar que en ningún caso hemos querido insultar al oyente que se considera tratado como un sirviente, etc. No puede, en general, exigírsenos responsabilidad por toda aquella información proposicional que a un oyente cualquiera pueda sin más sugerirle lo que decimos.

Hay un nivel de significado proposicional, sin embargo, con respecto al cual nuestro grado de responsabilidad parece ser crítico. Es el que está involucrado en la mayoría de nuestros intercambios destinados a objetivos concretos, el que puebla las revistas científicas, los periódicos,

etc. el que sirve habitualmente para que los demás expliquen y predigan nuestro comportamiento; el que nos puede llevar a la cárcel o arruinar nuestra carrera académica. Este nivel privilegiado de significado proposicional asociado con nuestras proferencias comunicativas es el nivel que Grice llamó 'lo que se dice'.

3. ¿Cómo individuar *lo que se dice*?

A partir de los trabajos más influyentes de Grice (1957/1989, 1975/1989) la noción de *lo que se dice* se ha entendido de muchas y diversas maneras. Con el paso del tiempo han ido surgiendo distintas posiciones que, al intentar explicar qué es *lo que se dice*, han dado preponderancia a distintos aspectos. Dentro de esta explicación ha jugado un papel predominante la forma en que se individúa *lo que se dice*. Ahora definiremos y explicaremos estas diversas estrategias de individuación y colocaremos la estrategia presentada por I. Stojanovic dentro de este marco plural de propuestas.

A) *Minimismo* (proposición mínima): *lo que se dice* está determinado por dos criterios: i) debe incluir la información necesaria y no más de la necesaria para obtener una entidad evaluable como verdadera o falsa, y ii) contiene solo la información demandada por los elementos pertenecientes a la forma lógica de la oración. Si existen elementos que no están lingüísticamente marcados pero que pueden hacerse covariar con la presencia de otros, entonces se consideran "ligados", es decir, pertenecientes a la forma lógica de la oración, y su interpretación será parte de la proposición expresada (ver Stanley 2000, 410, 412, 421; 2002, 150-152).

B) *Contextualismo* (criterio de disponibilidad): *lo que se dice* es una variedad de significado no-natural, es decir, es esencialmente una cuestión de expresión y reconocimiento de intenciones comunicativas. Para obtener las condiciones de verdad intuitivas de una proferencia dada no es necesario que los hablantes tengan la habilidad de hacer informes sobre lo que se ha dicho mediante dicha proferencia. Simplemente hay que situarlos delante de distintos escenarios y situaciones y pedirles que evalúen la proferencia en cuestión en relación a dichos escenarios y situaciones (ver Recanati 2004, 13-15). Este criterio es independiente de i) y suele ofrecer resultados incompatibles con ii).

C) *Minimismo del acuerdo*: el criterio de disponibilidad es interpretable como un criterio basado en el uso de los verbos "conocer" y "creer". *Lo que se dice* es lo que los hablantes competentes "creen" que dice una determinada oración cuando es proferida en un contexto. Cappelen y Hawthorne proponen para la determinación del contenido compartido, de *lo que se dice*, el "test del acuerdo", según el cual si podemos decir de dos hablantes que han llevado a cabo dos

proferencias distintas u' y u'' de la misma oración S que "estaban de acuerdo", entonces podemos considerar que *lo que se dice* es semánticamante invariante, que *lo dicho* mediante estas proferencias de S no cambia del contexto de u' al contexto de u''. (ver Cappelen y Hawthorne 2009, 54 y ss). Este criterio está en parte pensado para evitar la violación de i).

D) *I. Stojanovic*, *decir lo mismo*: si ante dos proferencias u' y u'' de dos hablantes A y B, un tercer hablante puede decir que "A y B dijeron lo mismo", entonces *lo que se dice* mediante las proferencias de u' y u'' contiene la información mínima común entre ambas, y es neutra con respecto al resto de la información asociada con ellas, ya sea contextual o lingüística. Este criterio no está comprometido con i), y en ese sentido es cercano al contextualismo, pero contiene la hipótesis de que *lo que se dice* pertenece al significado léxico, lo cual es cercano a la descripción de los mecanismos interpretativos a partir de la forma lógica propuestos por el minimismo, como hemos visto en ii).

4. Problemas para el criterio propuesto

La estrategia citada al principio del trabajo, basada en la determinación de *lo que se dice* a través de las declaraciones de cuándo *decimos lo mismo*, es crucial para fundamentar la idea de que *lo que se dice* es simplemente el significado léxico. Esta estrategia ocupa en la posición minimista de I Stojanovic el papel del Principio de Disponibilidad para las versiones de la Pragmática de las Condiciones de Verdad presentadas hasta 2007, y de las consideraciones relacionadas con las declaraciones de acuerdo en (Cappelen y Hawthorne 2009). Argumentaremos ahora que el hecho de que pueda decirse que dos proferencias dadas dicen lo mismo no es una condición suficiente para concluir que solo los elementos comunes a ambas proferencias forman parte de *lo que se dice*. Este criterio, como anunciamos más arriba, es demasiado *liberal*, pues permite incluir en el nivel de *lo que se dice* información que tiende a situarse fuera de este nivel. Consideremos el siguiente ejemplo:

(4) En el recreo siempre le pegan y le roban las chucherías (dicho por el Prof. Hernández).
(5) Pobrecillo, es la piñata del recreo (dicho por la Profa. Villa).
(6) Eso es lo mismo que dijo el Prof. Hernández, todos abusan de él (posible réplica a (5)).

De acuerdo con el criterio presentado en la primera sección de este trabajo, los elementos que formen parte de *lo que se dice* mediante las proferencias de (4) y (5) serán aquellos elementos

comunes a (4) y (5) que sean necesarios para poder afirmar que ambas proferencias dicen lo mismo (en la lectura más *flexible* posible). Si seleccionamos los elementos comunes a ambas proferencias que son necesarios para poder afirmar (6), nos vemos forzados a aceptar el significado metafórico de "piñata" como parte de la proposición expresada mediante (5). Si podemos afirmar que ambas proferencias dicen lo mismo es porque ambas dicen que en el recreo se abusa de un cierto individuo, ya sea mediante el significado literal de 'le pegan y le roban las chucherías' en (4), o mediante el significado metafórico de 'piñata' en (5). Esto produce resultados contra-intuitivos, ya que no es parte de una posición completamente estándar defender que el significado de una metáfora novedosa es parte de *lo que se dice*.

Los resultados contra-intuitivos producidos por el criterio no están restringidos al ámbito de las metáforas, sin embargo. Consideremos el siguiente ejemplo:

(8) O te comes los garbanzos o tendrás lo mismo para cenar (dicho por la madre de Pedro).

(9) Anda, justo lo mismo que me decía mi madre en todas las comidas (dicho por Juan como una posible réplica a (8)).

Si aplicamos de nuevo la estrategia de "Decir lo mismo", podemos ver que el elemento que nos permite decir que tanto la madre de Pedro como la madre de Juan dicen lo mismo es simplemente una cierta estructura sub-oracional formada por un condicional, un sujeto elidido y dos verbos transitivos sin CD determinado. Si aplicamos el criterio podemos inferir que *los garbanzos* no forman parte de *lo que se dice* en (8), ya que tenemos la intuición de que no es un elemento necesario para poder afirmar que ambas madres dicen lo mismo.

P*rima facie*, no diríamos que el sustantivo en función de complemento directo de un verbo transitivo no contribuye a *lo que se dice*. En este sentido, el criterio "Decir lo mismo" es demasiado *restrictivo*, puede excluir en ocasiones información normalmente considerada esencial para la determinación de *lo que se dice*. Los ejemplos suministrados para mostrar la insuficiencia del criterio "Decir lo mismo" apuntan en una misma dirección: tenemos la intuición de que podemos decir que dos proferencias han dicho lo mismo en una gama amplia de casos que van desde proferir las mismas palabras, sean o no oraciones completas, hasta expresar la misma proposición (usemos las mismas palabras o no).

Conclusión

Las diferencias que vimos en la sección 4 entre los distintos criterios para individuar *lo que*

se dice pueden abordarse desde distintas perspectivas. Puede considerarse, como hemos visto, qué tipo de actitud proposicional o verbo de informe se usa para elaborar el criterio (el contextualismo usa los verbos "decir" y "creer", y el minimismo del acuerdo los verbos "estar de acuerdo" y "estar en desacuerdo", I Stojanovic la locución "decir lo mismo", etc.). Puede tenerse en cuenta también si la información que incluimos en *lo que se dice* está demandada lingüísticamente o no, o si es necesaria para obtener un nivel de contenido evaluable como verdadero o falso (compromiso del minimismo con la proposición mínima). Además de estas consideraciones, puede también considerarse el papel que las intuiciones de los hablantes acerca de la verdad de lo que dicen juegan en cada marco teórico. Solo el contextualismo ofrece un punto de partida desde el que se puede obtener una posición no revisionista acerca de *lo que se dice*: las intuiciones de los hablantes son los datos brutos que hay que acomodar teóricamente, no una fuente de información revisable desde la teoría. Para las distintas variantes del minimismo, sin embargo, las intuiciones de los hablantes son claramente revisables; muchas veces no "vemos" el contenido semántico de nuestras propias proferencias. Considérese al respecto esta declaración de intenciones de Cappelen y Hawthorne: "En la medida en que a veces somos ciegos a las diferencias en el contenido semántico, seremos propensos a hacer una mala aplicación de los tests. De este modo los diagnóticos de acuerdo no nos protegen de hacer malos juicios sobre la uniformidad y las diferencias de contenido, ya que los juicios defectuosos sobre dicha uniformidad y diferencias producirán juicios defectuosos sobre qué informes de acuerdo o desacuerdo son correctos" (Cappelen y Hawthorne 2009, 66-67).

Todos estos criterios son intentos de individuar un nivel de significado privilegiado (*lo que se dice*), pero afrontan esta tarea desde presupuestos totalmente dispares. El criterio aquí analizado tiene, frente a otros, la enorme ventaja de ser testable. Poner a prueba el mismo nos ha llevado a concluir que es demasiado *liberal* y demasiado *restrictivo*. Inadecuado por tanto con respecto a nuestras intuiciones acerca de *lo que se dice*. La medida en la que estas intuiciones acerca de la insuficiencia del criterio se deban a nuestra ceguera con respecto a las propiedades semánticas de los ejemplos propuestas nos resulta, obviamente, inapreciable.

REFERENCIAS

Cappelen, H. & Hawthorne, J. 2009, *Relativism and monadic truth*. Oxford, Oxford University Press.
Grice, H. P. 1957, Meaning. *Philosophical Review,* 66**,** 377-388.
Grice, H. P. 1975, Logic and conversation. *Syntax and semantics,* 3: Speech acts**,** 41-58.
Recanati, F. 2001, What is said. *Synthese,* 128**,** 75-91.
Recanati, F. 2004, *Literal meaning,* Cambridge, Cambridge University Press.
Stanley, J. 2000, Context and logical form. *Linguistics and Philosophy,* 23**,** 391-434.

Stanley, J. 2002, Making it articulated. *Mind & Language,* 17**,** 149-168.
Stojanovic, I. 2008. *What Is Said: an Inquiry into Reference, Meaning, and Content.* Saarbrucken: VDM Verlag.

LUCK AND FORTUNE, OR WHY KNOWLEDGE REQUIRES COGNITIVE CONTROL

FERNANDO BRONCANO-BERROCAL
Universitat de Girona – LOGOS
ferbroncano@gmail.com

ABSTRACT: In this paper I give a definition of the notion of fortune and I distinguish it from the notion of luck. In the light of this distinction, I identify a knowledge-undermining phenomenon that has not been identified as such and which is different from epistemic luck: epistemic fortune. I advance the hypothesis that most (if not all) not known true beliefs are cases either of epistemic luck or epistemic fortune and I sketch a general strategy of dealing with both knowledge-undermining phenomena at once.
KEYWORDS: *epistemic luck*, *epistemic fortune*, *knowledge*, *control*

Although there has recently been an increasing body of literature on the notion of luck, not much has been said about the notion of fortune. In particular, nothing has been said on the implications it may have for knowledge. This short paper attempts to shed some light on the topic. The aims are four: 1) To draw the distinction between the notions of luck and fortune and to give a proper account of the latter. 2) To identify a knowledge-undermining phenomenon that, hitherto, has not been identified as such: veritic epistemic fortune. 3) To advance a hypothesis: most (if not all) cases of true belief that does not amount to knowledge are cases either of dangerous epistemic luck, or of dangerous epistemic fortune. 4) To briefly sketch the strategy that a theory of knowledge could follow in order to rule out at once cases of epistemic luck and epistemic fortune.

Let me introduce the basic intuitions. Compare these two cases. Lottery A is fair. S, believing that the lottery is fair, buys a ticket and wins. Lottery B is rigged in S's favor. S, believing that the lottery is fair, buys a ticket and wins. The intuition that everyone will have concerning Lottery A is that S is lucky to have won. After all, fair lottery wins are the paradigmatic cases of luck. However, is S lucky to have won Lottery B? People's intuitions with respect to Lottery B can be divided into two groups. People in the first group straightforwardly deny that S is lucky to have won the lottery. When asked why, they will claim that S cannot be lucky because someone has decided that she would win. By contrast, people of the second group are willing to claim that S is lucky to have won. When asked why, they will explain that, even though Lottery B was rigged in

her favor, S did not know it and, in this way, she is lucky that a secret benefactor decided to give her the prize.

I will present a plausible error theory that explains why people in the second group have the wrong intuition. The first thing we should note is that there are important structural differences between the two lotteries. Before winning Lottery A, there was very low objective probability that S would win (suppose that the lottery drum contains a huge number of balls). By contrast, before winning Lottery B, there was high probability (near 1) that S would win (here it is irrelevant the number of balls in the drum: the secret benefactor manipulates the lottery system anyway).

Let us state the difference once again, but this time in modal terms. In most close possible worlds where S buys a ticket of Lottery A, she does not win, while in most close possible worlds where S buys a ticket of Lottery B, she does win. Close possible worlds are obtained by making a small change to the actual situation just when the relevant event is about to occur (or shortly before). For example, with respect to Lottery A, possible worlds that count as close are worlds in which the balls in the lottery drum follow a different trajectory. With respect to Lottery B, possible worlds that count as close are worlds in which the secret benefactor still rigs the lottery in S's favor (we can give more support to this point by adding some extra details: the secret benefactor's decision to rig the lottery in S's favor is a firm decision, she is the owner of the lottery, she knows how to manipulate a lottery system, and so on and so forth). In short, close possible worlds are worlds very similar to the actual one.

Some commentators (Coffman 2007: Levy 2009, 2011; Pritchard 2005) take the following to be a necessary condition for luck: an event E is lucky only if E occurs in the actual world but does not in (most) close possible worlds where the relevant initial conditions for the occurrence of E are the same (or very similar) to those in the actual world (the relevant initial conditions for a lottery win is, let say, that one buys a lottery ticket). In short, lucky events are modally fragile events.[1]

Winning Lottery B is not lucky in this sense because S does win the lottery in close possible worlds in which she buys a ticket. Why it is then that people in the second group have the intuition that she is lucky? Here is the error theory: because they are conflating luck with fortune. S is fortunate but not lucky to win Lottery B. People in the second group have wrong intuitions because fortune and luck are very close notions, but there is a fundamental difference between them: an event E is fortunate only if E occurs in the actual world and in close possible ones where the

1

 In addition, in order to be lucky the relevant event must be significant for the agent, since there are modally fragile events that are not lucky (e.g. a fortuitous rockfall that affects no one).

relevant initial conditions are the same.[2] Note that this condition is not sufficient for fortune since there are lots of events that are modally robust that do not count as fortunate (e.g. going to work every day).

There is another condition shared both by luck and fortune. Several commentators (Coffman 2007; 2009; Latus 2003; Levy 2009; Riggs 2007) plausibly, though not undisputedly, think that in order for an event E to count as lucky for S, S must lack control over E (alternatively: E must be significantly beyond S's control). For example, a fair lottery win is lucky because the winner has no control over the outcome. If S rigged the lottery, S would gain control over the outcome and thus it would not be by luck that she would win. Coffman (2007; 2009), Latus (2003), Levy (2009) concatenate the lack of control condition with the modal condition to provide hybrid definitions of luck. My point is that we can do the same in the case of fortune:

> **Fortune**: E (a significant event for S) is fortunate for S if and only if E occurs in the actual world and in close possible worlds where the relevant initial conditions for E are the same as in the actual world, and S lacks control over the occurrence of E.[3]

S is fortunate to have won a rigged lottery because, although her win is modally robust, S lacks control over the outcome of the lottery (unbeknownst to S, someone else rigged the lottery). In the same way, it is good fortune (but not good luck) that the sun rises every morning, since sunrises are very modally robust events over which we lack control. Another (actual) example: it is not by luck that we are in a long-term crisis; rather, it is bad fortune for us, citizens, as the development of this crisis is something beyond our control and (presumably) a modally robust phenomenon (in virtue of the cyclical structure of capitalism, the sustained economic interests of some economic and political groups, the systematically disastrous economic policies of the different governments, etc.).

2

 This condition can be found, more implicitly than explicitly, in some responses by Coffman (2007) and Levy (2009) to counterexamples to the previously stated modal condition for luck. In addition, as in the case of luck, in order for an event to count as fortunate, the event must be significant for the agent. In what follows, I will take for granted the necessity of a significance condition both for luck and fortune.

3

 As it is commonly stipulated with respect to luck, I use the term 'fortunate' to refer both to instances of bad and good fortunate.

In conclusion, while both lucky and fortunate events are significant events over which we lack control, the difference between both is that lucky events are modally fragile while fortunate ones are modally robust.

Let us turn on how the distinction bears on knowledge. The exclusion of cases of dangerous epistemic luck has been one of the main concerns of contemporary epistemologists since Gettier's challenge. Thanks to recent work (mainly by Pritchard 2005) we have now a very good understanding of the phenomenon. The kind of knowledge-undermining luck involved in Gettier-style cases is veritic luck, where a veritically lucky belief is a belief that hits the truth by luck. Applying a hybrid definition of luck, S's belief that p is luckily true if and only if S has no control over the hitting on truth of her belief and in most close possible worlds where S believes that p, p is false.[4] In this way, gettiered beliefs are modally fragile true beliefs the hitting on truth of which is not controlled by the agents who have them.

Note that the relevant cognitive control at issue is not voluntary control over belief formation (neither direct, nor even indirect). Indirect doxastic voluntarism may be true but direct doxastic voluntarism is surely false and we do not want the problems it has. Rather, the kind of cognitive control at issue is control (I think mainly involuntary but not necessarily) over the hitting the truth of the relevant belief and not over its formation, i.e. it is control over the way an involuntary formed belief gets its truth value. I do not intend to give here a full account of the kind of cognitive control that is required to exclude knowledge-undermining luck. The account of this type of control, which I will call 'veritic cognitive control', will hinge dramatically on our intuitions about knowledge (e.g. on whether we are externalists or internalists). The important thing that we must keep in mind that there is a sharp distinction between veritic cognitive control and the kind of doxastic control entailed by doxastic voluntarism.

The main reason to introduce the notion of veritic cognitive control is that it allows us to identify a knowledge-undermining phenomenon that threatens true belief and that is different from knowledge-undermining luck. There are cases of true beliefs that, despite being modally robust (i.e. true in most close possible worlds), do not amount to knowledge because the agents in question lack

4

 Perhaps also: in most close possible worlds where S believes a proposition closely related to p in an epistemically relevant way, the proposition is false.

veritic cognitive control over their hitting on truth. These are cases of dangerous epistemic fortune, or so I argue. Here it is its most extended variety:[5]

> **Veritic Epistemic Fortune (VEF)**: S's belief that p is true by fortune if and only if S has no control over the hitting on truth of her belief and in most close possible worlds where S believes that p, p is true.

There are many cases of not-known true belief that are neither Gettier-style cases, in particular, nor cases of veritic luck, in general, but which fall under the definition of VEF. A case of VEF for example is a case in which S forms by wishful thinking the false belief that E will occur and a benevolent demon changes the future so that S's belief turns out to be true. This other case, given by Pritchard (*forthcoming*) against the safety principle, is also a case of VEF: S forms beliefs about the temperature in her room by consulting a thermometer on the wall, which, unbeknownst to S, is broken and fluctuating randomly within a given range. But there is a hidden agent manipulating the temperature of the room in such a way that she adjusts the temperature of the room every time S looks at the thermometer so as to fit the readings S sees. In this way, S never forms a false belief. In both cases the beliefs in question are modally robust, but the agents seem to lack the kind of cognitive control over the hitting on truth that is required for knowledge.

The identification of veritic epistemic fortune as a proper knowledge-undermining phenomenon allows us to state the following hypothesis, which I consider very plausible: most (if not all) cases of the literature of true belief that does not amount to knowledge are cases of either dangerous epistemic luck, or of dangerous epistemic fortune (e.g. VEF). A quick look at the cases of the literature gives support, at least *prima facie*, to the hypothesis.

In any case, the point is that if we take an abstract point of view of the situation in the face of that hypothesis, we can delineate the general strategy that a theory of knowledge could follow to rule out at once cases of veritic epistemic luck and veritic epistemic fortune. Here is the guideline. Since lacking veritic cognitive control is a necessary condition both for veritic luck and veritic fortune, what a theory of knowledge needs to do in order to rule out at once most (if not all) cases of not-known true belief is to develop a notion of veritic cognitive control that implies that if an agent

5

 As in the case of luck, I do not want to exclude neither the possibility of there being more dangerous types of epistemic fortune, nor the possibility of there being forms of epistemic fortune compatible with knowledge.

has cognitive control over the hitting truth of her belief that p, then p is true in most close possible worlds (i.e. modally robust).

Let us take stock. The aim of this paper has been simply to identify a knowledge-undermining phenomenon different from epistemic luck, namely veritic epistemic fortune, and to explain the implications it may have for our philosophical theorizing in epistemology. The aim has not been to account for the (until yet) undefined notion of veritic cognitive control (and to give, accordingly, an account of knowledge), in part because, as I said, how we conceive this notion will crucially depend on what we think necessary for knowledge. Nevertheless, I would like to conclude by briefly presenting some possible ways of implementing the notion and the strategy sketched above.

From an externalist point of view, it seems that an agent manifests veritic cognitive control over the hitting on truth of her belief that p when the agent forms the belief that p in a reliable way. For example, S has veritic cognitive control over the hitting on truth of her belief that there is hand before her, only if S forms the belief by using a reliable method (e.g. a reliable cognitive process like vision). Moreover, I suspect that an externalist should go for the strongest claim that the relevant cognitive process must be cognitively integrated, especially in the face of cases like the following: S has a brain lesion that causes a reliable cognitive process in his brain him that produces the belief that he has a brain lesion (the case is by Plantinga 1993). The idea is that unless our belief is formed by a cognitively integrated reliable process we cannot be said to exert control over the fact that it hits the truth. This is the line that virtue epistemologists would surely follow (a reliable belief-forming process count as epistemically virtuous only if cognitively integrated).

On the other hand, some other cases might lead us to a more internalist position concerning veritic cognitive control. It seems that in some cases cognitive control over the hitting on truth of a belief can be manifested when the belief-forming method is unreliable. For instance, if the agent acquires background knowledge of the conditions under which her belief gets the truth value. For example, imagine that S forms the false belief, by wishful thinking, that she will win the lottery. If before the lottery takes place, she comes to know that the lottery is rigged in favor of someone else, she is a position to rectify her (irrational) belief that she will win (in this sense, knowledge compensates the unreliability of the method). In the light of this case, some internalists would say that veritic cognitive control requires direct access to the reasons on the basis of which the belief has been formed. By contrast, some virtue epistemologist would keep the notion externalist and perhaps would simply claim that veritic cognitive control is sometimes exerted at the level of

reflective knowledge.[6] But I will leave the question of how to account for the notion of veritic cognitive control open. For my aims have been to account for the notions of fortune and veritic fortune and to show that in the face of veritic fortune and veritic luck, the best that our theory of knowledge can do is to follow a unitary strategy based on the notion of veritic cognitive control, whatever it amounts to.

REFERENCES

Coffman, E. J. 2007, 'Thinking About Luck' *Synthese* 158 (3): 385 – 398.
———. 2009. 'Does Luck Exclude Control?' *Australasian Journal of Philosophy* 87 (3): 499–504.
Latus, A. 2003, 'Constitutive Luck' *Metaphilosophy* 34 (4): 460–475.
Levy, N. 2009, 'What, and Where, Luck Is: A Response to Jennifer Lackey' *Australasian Journal of Philosophy* 87 (3): 489 – 497.
Plantinga, A. 1993. *Warrant and Proper Function*. Oxford University Press.
Pritchard, D. *forthcoming*, 'Anti-Luck Virtue Epistemology' *Journal of Philosophy*
———. 2005. *Epistemic Luck*. Clarendon Press.
Riggs, W. 2007, 'Why Epistemologists Are so down on Their Luck' *Synthese* 158 (3): 329 – 344.
Sosa, E. 2011, *Knowing Full Well*. Princeton University Press.

6

Sosa's theory of knowledge (2011) seems to be compatible with this line of response, since it is crucial to his theory the distinction between 'animal' (first-order) and reflective (second-order) knowledge. According to him, knowing full well requires both.

IT IS NOT THE CASE THAT THE ALLEGED GOLDEN MOUNTAIN EXISTS[1]

GEMMA CELESTINO FERNÁNDEZ
University of British Columbia,
Universitat de Barcelona & LOGOS
gceles@interchange.ubc.ca

ABSTRACT: In this paper I will discuss a recent proposal by Frederick Kroon (2004, 2009). It belongs to a broad family of views, according to which sentences containing empty names do not even semantically express any proposition, but we use them, and pretend them to be true, in participating in games of make-believe. This is in a nutshell the proposal of Gareth Evans (1982) and Kendall Walton (1990). I will first discuss Kroon's positive view and then I will turn to his criticism to what he calls 'the Ellipsis Strategy' to argue that his arguments are not successful.

According to Kroon, by means of pretense, speakers who utter negative existentials such as (1)

(1) E.T. does not exist

manage to assert something informative, even if these negative existentials are otherwise contradictory. In general, when speakers utter contradictory sentences such as (1), they pretend that they refer to something and then they use this pretense to finally assert that there is no such individual. On this view, what speakers assert by uttering such sentences is not what these sentences would literally mean. It is not semantically, but pragmatically, expressed.

For instance, in uttering (1) speakers pretend that they refer to some individual that exists by their use of the name 'E.T.', and use the resulting interpretative tension to assert, in part, something as follows:

(2) Outside of the pretense that there is a unique actual individual called 'E.T.' whom I have seen or heard described as a homesick alien who wants to go back home, etc., there is no such individual.

This is how Kroon *allegedly* solves the problem of true negative existentials.

[1] I would like to thank Pablo Rychter and Ori Simchen for helpful discussions and comments on previous versions of this paper. Research leading to this work has been partially funded by the research project FFI2011-29560-C02-01 (Spanish Government), and the consolidated and funded LOGOS research group 2009 SGR 1077 (Catalan Government).

Let me very briefly mention several worries that I have about this account. First, in my opinion, this proposal does not make sufficiently clear how these contents, that according to it we communicate in uttering sentences with empty names, are pragmatically conveyed. Kroon rejects a Gricean explanation of them. Second, the account explains all content communicated by sentences with empty names as pragmatically conveyed. That is, we never use these sentences to communicate their literal content. And I find this suspicious for at least the following two reasons: one reason is that it strikes me as odd that no utterances of sentences in which empty names occur, which are nevertheless heavily used, have literal content. This would mean that we use sentences without meaning a lot, while understanding each other without problems nonetheless. The other reason is that it is true that pragmatics plays a major role in communication, but it also seems true to me that pragmatics crucially rests on semantics and the literal contents of the sentences uttered. The Gricean explanation of conversational implicatures is a paradigmatic case of a plausible pragmatic explanation, but it crucially depends on the semantics in the sense that it starts from the semantic content literally expressed by utterances of sentences. Given that according to pretense-views like Kroon's, utterances of sentences containing empty names do not semantically express any proposition, this kind of pragmatic explanations for the other propositions (which according to it are pragmatically conveyed) is not available to them. Third, the pragmatic content that according to Kroon's proposal gets communicated is explicitly about pretense, when it does not seem that people uttering these sentences are saying anything about it –even if it is true that they are fully aware that they are talking about fiction and they often engage in pretense when discussing fiction. Fourth, the proposal does not make a distinction between the fictional case and other cases that involve non-fictional empty names such as 'Vulcan', in which it is not even clear that there is ever any pretense involved.

Having outlined and briefly discussed Kroon's pragmatic proposal, let us now focus on an argument that Kroon advances for his view -an argument that in fact has a much wider interest and motivation. Kroon argues that the typical Russellian treatment of negative existentials that contain definite descriptions does not work. Consider

(3) The golden mountain –you know, the elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration- does not in fact exist.

On the typical Russellian treatment, this is understood with negation taking wide scope over the description. However, against this widespread treatment, Kroon interprets (3) with negation taking narrow scope instead.

Kroon argues as follows. The following cannot be (3)'s interpretation

(4) The entity that is at once the golden mountain and the elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration does not exist.

interpreted, in its turn, as

(5) It is not the case that there is a unique entity that is at once the golden mountain and the elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration.

because (3), but not (4) or (5), is false in some worlds in which there is no golden mountain and Dr. Zak did not write any book about journeys of exploration.

Kroon further argues that (6) cannot be an interpretation of (3) either, because it requires both the existence and the non-existence of the golden mountain.

(6) The golden mountain does not exist, and (the golden mountain = the elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration)

To see this notice that on the typical Russellian treatment, the first conjunct of (6) would be interpreted as

(7) It is not the case that there is a unique golden mountain

and the second conjunct, as

(8) There is a unique golden mountain and there is a unique elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration and they are the same thing.

Thus, on the typical Russellian treatment, (7) asserts the non-existence of the golden mountain while (8) asserts the existence of the golden mountain. Therefore, on the typical Russellian interpretation, (6) cannot be true, while (3) can.

According to Kroon, the use of definite descriptions in negative existentials "typically *feels* as if it brings with it a commitment to there being an object of a certain sort, even as the speaker denies that there is such an object." (Kroon 2009, 374-375). And he claims to be able to account for this by appealing to pretense.

Like in the cases of true negative existentials involving empty names, Kroon proposes to explain this as follows: utterances of (3) semantically express something inconsistent, such as

(9) There is a unique golden mountain that does not exist.

but (by means of pretense) manage to pragmatically express and assert something which is not, such as

(10) Dr. Zak used the description 'the golden mountain' in his book on famous journeys of exploration and no one has ever discovered such a mountain. It is not the case that there is a unique golden mountain.

Against what is commonly accepted, Kroon claims that in general, true negative existentials involving definite descriptions should be interpreted with negation taking narrow scope under the existential quantifier.

However, Kroon's interpretations should be resisted. It seems that the examples that he discusses to make his point admit of other readings compatible with the usual Russellian treatment of them, according to which negation takes wide scope over the existential quantifier. These are cases in which something like 'alleged' is elided as part of the definite descriptions involved. Thus, for instance, I would treat (3) as

(11) The *alleged* golden mountain –you know, the *alleged* elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration- does not in fact exist.

interpreted as

(12) It is not the case that there is a unique entity that is both the alleged golden mountain and the alleged elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration.

Unlike (4) and (5), but like (3), (11) and (12) would be false in the world considered before and would not be inconsistent like (6). (11) and (12) would be true only in those possible worlds in which at least someone thought of or pretended that an alleged golden mountain of a certain sort existed; in particular, an alleged golden mountain that had allegedly also been an elusive mountain that Dr. Zak wrote about in his book on famous journeys of exploration. Further, it would also account for the pretense phenomenology that, as Kroon points out, accompanies these statements. Thus, they would be good interpretations of the complex existential (3).

The explanation that I am offering, which appeals to the existence of an elided expression such as 'alleged', is a version of what Kroon calls 'the Ellipsis strategy'. Kroon criticizes this strategy, arguing that it does not provide correct interpretations. But it does. Let us now see this with the following example that Kroon also asks us to consider.

(13) The "splendid and immaculate" hotel you booked me into was neither immaculate nor splendid.

I propose to interpret this as follows.

(14) The *allegedly* splendid and immaculate hotel you booked me into was neither immaculate nor splendid.

, which, in turn, is interpreted as follows

(15) There is a unique allegedly splendid and immaculate hotel you booked me into that was neither immaculate nor splendid.

It should be noticed that the specific interpretations of the ellipsis strategy that Kroon objects to are different to the ones I propose. In the case of (13), for instance, Kroon objects to the following.

(16) The hotel you booked me into, which you described as 'splendid and immaculate', was neither immaculate nor splendid.

According to Kroon, there is no syntactic or semantic evidence in favor of such a strategy, but there is much evidence against it. His objections are the following. He first adduces the fact that there is no explicit occurrence of a phrase like 'described as' and then objects that it is implausible to think that the expression 'splendid and immaculate' is being mentioned rather than used, even if it appears within quotation marks, because the speaker could have used finger quotes, a certain tone of voice, or even the recognition by her audience of her expectation to be understood as speaking from another's perspective when uttering this expression, rather than used these quotation marks, and she would have managed to assert the same.

However, these objections are wrong. First, one often utters sentences such as (16), (14) or (15), instead of (13), making the elided phrase explicit seemingly to make the same point. Secondly, the fact that a sentence such as (13) can be uttered without any kind of quotation marks, or even without any tone of voice, does not show that the expression that appears within quotation marks in (13) is not being somehow mentioned or used as part of a demonstration of something, and that the claim that one makes by uttering (13) is not made by precisely these means. In fact, finger quotes, a tone of voice, or even speakers' intentions might perfectly play the role that written quotation marks play in paradigmatic cases of quotation.

Kroon also objects that the readings proposed by the ellipsis strategy that he considers, in which it is explicitly said that the hotel has been described as 'splendid and immaculate', miss the irony that exists in utterances of sentences such as (13). According to Kroon, this irony is even more clearly present in sentences such as (17).

(17) The "splendid and immaculate" hotel you booked me into –you know, the one where maids vacuum the floors every hour and the guests are mostly Saudi princes- was neither splendid nor immaculate.

But Kroon is wrong in this respect as well: strong irony or mimicking may be present even if the point is made more explicitly by explicitly uttering phrases such as 'that you described as' like in

(18) The hotel you booked me into which you described as 'splendid and immaculate' –you know, the one that according to a description like yours would have maids vacuum the floors every

hour and guests would be mostly Saudi princes- was neither splendid nor immaculate.
or phrases such as 'allegedly' like in

(19) The allegedly splendid and immaculate hotel you booked me into –you know, the one where allegedly maids vacuum the floors every hour and the guests are mostly Saudi princes- was neither splendid nor immaculate.

That is, irony is something extra to what is said: some claims, even (18) and (19), might be made with irony or without it – even if it is true that irony works better if certain things are not made explicit. This might be precisely why some tend to utter sentences like (13) or (17) rather than sentences like (14)-(16), (18) or (19).

The same goes for the other utterances that Kroon considers when making this objection, such as

(20) Do you have any other "splendid and immaculate" hotels to show me?
and

(21) You know the "splendid and immaculate" hotel you booked me into? Amazingly enough, and contrary to all the logic I know, it was neither immaculate nor splendid.
, which according to the ellipsis strategies would be interpreted as

(22) Do you have any other hotels that you would describe as 'splendid and immaculate' to show me?
or as

(23) Do you have any other allegedly splendid and immaculate hotels to show me?
and as

(24) You know the hotel you booked me into which you described as 'splendid and immaculate'? Amazingly enough, and contrary to all the logic I know, it was neither immaculate nor splendid.
or as

(25) You know the allegedly splendid and immaculate hotel you booked me into? Amazingly enough, and contrary to all the logic I know, it was neither immaculate nor splendid.

Finally, Kroon objects that the ellipsis strategy would render false the otherwise intuitively true Russellian interpretation of negative existentials such as

(26) The "splendid and immaculate" hotel that you booked me into does not exist –the rooms are a disgrace.

that arise out of a restrictive reading of the embedded relative clause 'which is "splendid and immaculate"' of sentences such as (13). And Kroon seems to think that the true reading is

(27) It is not the case that there is a unique hotel you booked me into, which was "splendid and immaculate" –the rooms are a disgrace.

and that the Russellian interpretation on the ellipsis strategy, though, is

(28) It is not the case that there is a unique hotel you booked me into that you described as 'splendid and immaculate' –the rooms are a disgrace.

which is false, given that he did book her into a hotel which he described as 'splendid and immaculate'.

On my version of the ellipsis strategy, (26)'s Russellian interpretation would turn out to be as follows instead.

(29) It is not the case that there is a unique allegedly splendid and immaculate hotel you booked me into –the rooms are a disgrace.

which in turn Kroon would take to be interpreted as the false

(30) It is not the case that there is a unique hotel you booked me into which was allegedly splendid and immaculate –the rooms are a disgrace.

However, the ellipsis strategies are not forced to take these false readings of the Russellian interpretation of (26). Contrarily, the right readings seem respectively to be instead

(31) It is not the case that there is a unique hotel you booked me into which was splendid and immaculate as you described it to be –the rooms are a disgrace.

and

(32) It is not the case that there is a unique hotel you booked me into which was splendid and immaculate as it was allegedly to be –the rooms are a disgrace.

which are true, rather than false as objected. There is nothing that prevents these true readings of

(26) on the ellipsis strategy and so Kroon's final objection does not work either.

Kroon then considers negative existentials whose truth does not involve any object at all, such as the complex and long

(33) The fabulous object that Jones discovered –the golden mountain, no less, as is so clearly shown (so Jones keeps reminding us) by the golden streaks on the last slide- is, sad to say, neither golden nor a mountain. Look closely enough, and you'll see that it does not even exist. Jones's fabulous object is just a trick of the light –there is nothing there.

uttered with sarcasm.

According to Kroon, the ellipsis strategy that he considers cannot explain (33) either. Unlike the case of the hotel, in that other case there is not even a thing to be described. However, note that even if the version that Kroon considers does not work, my version of the ellipsis strategy does

work in that case too, and so contrarily to what Kroon assumes, there is an ellipsis strategy that works. On this version, (33) is taken to be an abbreviation of

(34) The *alleged(ly)* fabulous object that Jones *allegedly* discovered –the *alleged(ly)* golden mountain, no less, as is *allegedly* so clearly shown (so Jones keeps reminding us) by the golden streaks on the last slide- is, sad to say, neither golden nor a mountain. Look closely enough, and you'll see that it does not even exist. Jones's *alleged(ly)* fabulous object is just a trick of the light – there is nothing there.

Therefore, my proposed version of the ellipsis strategy gives a simpler and unified account of all cases, and at the same time agrees with the widely hold thought that Russell was right on negative existentials involving descriptions.

Summing up, in this paper I have discussed Kroon's Pragmatic Pretense-Based account of sentences with empty names, and his arguments against the typical Russellian treatment of negative existentials involving definite descriptions and the ellipsis strategy that I favor. If I am right, Kroon's account and arguments are wrong.

REFERENCES

Evans, G. 1982, *Varieties of Reference*. OUP.
Kroon, F. 2009, 'Existence in the Theory of Definite Descriptions', *Journal of Philosophy* 106 (7), 365-389.
----- 2004, 'Descriptivism, Pretense and the Frege-Russell Problems', *The Philosophical Review* 113 (1), 1-30.
Russell, B. 1905, 'On Denoting', *Mind* 14 (56), 479-493.
Walton, K. L. 1990, *Mimesis and Make-Believe: On the Foundations of the Representational Arts*. HUP.

COMMUNICATIVE INTENTIONS AND JOINT COMMITMENTS IN SPEECH ACTS

CRISTINA CORREDOR
University of Valladolid
corredor@fyl.uva.es

ABSTRACT: The aim of this contribution is to examine the notion of commitment involved in some alternative views of the illocutionary force of speech acts. After examining the Gricean approach, where a notion of commitment can be said to be connected to the expression of attitudes on the part of a speaker, the interactionalist approach is taken into consideration to highlight those aspects of the interaction where a *joint* commitment is required for the illocutionary act to take place, thus going beyond the speaker's communicative intentions. Finally, some difficulties in this last account will be taken into consideration.

KEY WORDS: illocutionary acts, commitment, communicative intentions, joint meaning

1. Commitment and speech acts

The aim of this contribution is to examine the notion of *commitment* involved in some alternative views of the illocutionary force of speech acts. To the extent that such a normative notion is required to account for communication and speech acts, the naturalistic import of a Gricean approach seems questioned. It has been argued, however, that communicative intentions and certain relevant normative attitudes can be connected to each other, in such a way as to make the latter a necessary condition of the former (Harnish 2005). On the other side, interactional approaches have highlighted the active role played by the hearer in determining the final force of a speech act, contending that communication should be seen as a joint project where participants jointly commit to the fact that a specific speech act has been performed (Carassa & Colombetti 2009). Ultimately, what is at stake here is the possible integration of both approaches in a more comprehensive view of speech acts.

The common conceptual framework is classical speech act theory, as initially formulated by Austin (1962) and later on developed by Searle (1969). A main thesis of the theory is that communication can be analyzed in terms of speech acts, and these in their turn as consisting of a locutionary dimension (belonging to the semantic level) and an illocutionary dimension (the pragmatic force of the speech act, its mode of communication, on which the action performed *in*

saying depends). Many theoretical efforts have been dedicated to the aim of clarifying what constitutes the illocutionary force of an utterance.

2. Expression of attitudes and responsibility: the Gricean approach

A very influential answer to that question broadens Grice's notion of *speaker's meaning* to further explain the force of an utterance in terms of specific speaker's communicative intentions. These are reflexive complex intentions that consist of the intention to express an attitude to an audience, intending to have this intention recognized, and intending that this recognition be part of the reason that the audience has to believe in the speaker's having this attitude.

In confrontation with this intentionalist account, it has been argued that the illocutionary force of an utterance is essentially dependent on a certain kind of commitment. Authors have contended that, in the particular case of an assertion that *P* (in what follows, the present discussion will focus on this case), the force depends on a commitment to the truth that *P* (Searle 1969), or that it depends on taking responsibility for the truth of *P* (Alston 2000). In confrontation with these objections, Harnish has recently stated that the following holds (Harnish 2005, 35):

(E ↔ C) In uttering *e*, *S* is expressing the belief that *P* ↔ In uttering *e*, *S* takes responsibility for its being the case that *P*.

In order to justify claim (E→ C), he develops the following analysis in two steps: if the speaker's expressing the belief that *P* (in uttering *e*) involves giving the hearer reason to think that the speaker holds *P* true, then one takes responsibility for those reasons being reasons for the truth of *P*. Furthermore, if the speaker takes responsibility for something being a reason for the truth of *P*, then one takes responsibility for the truth of *P*, in the absence of overriding reasons to the contrary (Harnish 2005, 36-37). The notions of undertaking responsibility and commitment are considered to be germane here (p. 23), hence no loss of generality is brought about by limiting the analysis, for the sake of simplicity, to the first one.

Two critical remarks seem appropriate here. Firstly, the inference goes through from the left to the right hand of the biconditional, but Harnish does not justify in detail the other direction of the arrow. The right-to-left inference is related to Moore's paradox and to the plausibility of a speaker undertaking responsibility for the truth of *P* without believing it. Previously to the formulation of the biconditional, Harnish has distinguished an achievement and a non-achievement use of "express", and has stated that "[l]ogically, expressing an attitude on the achievement use entails having the attitude" (p. 17), since this is the case in which the speaker makes "outer", overt or

manifest something that is effectively "inner". Notwithstanding this, it should be noticed that one thing is to hold a belief (and make it manifest), and another to undertake the responsibility of an assertion, in the sense of being in a situation to provide some justification for it.

In a subsequent essay, Harnish has introduced a distinction relevant here between success vs. non-defectiveness conditions of a communicative illocutionary act: the act is successful when the speaker has the appropriate intention, whereas it is non-defective when this attitude is satisfied (Harnish 2009, 20). This line of discussion cannot be prosecuted here, although it should be noted that the correction of the inference depends on the type of approach undertaken in the analysis. (E.g., a rule-epistemic approach might yield a different result from an inferentialist one. Harnish declaration that the connection is "more direct" than rule-epistemic seems to leave the problem unsolved).

Secondly, and concerning the first step of the analysis, the "reason" mentioned in the antecedent clause is very different in nature from the "reasons" referred to in the consequent. The first is a Gricean reason given to the hearer for him or her to attribute a propositional attitude to the speaker. The "reasons" in the consequent clause are those that can give support to the assertion and the corresponding belief. A possible interpretation of Harnish's formulation is this: if the speaker's utterance *e* provides the hearer with a reason to believe that the speaker has the relevant attitude (here a Gricean complex attitude that comprises the belief that *P*), then the speaker's utterance *e* provides the hearer with a reason to believe that "those reasons [that are] reasons" for the truth of *P* are available to the speaker.

The movement from belief to *justified* belief accomplished here could be seen as a requirement of rationality that underlies the illocutionary act of assertion, thus plausibly generating a responsibility or commitment from the part of the speaker to its being the case that *P* and a corresponding legitimate expectation in the hearer. This connection gives further support to Harnish's interesting remark that the Gricean notion of expressing an attitude is liable to criticism and blame, in the same sense in which a commitment to truth and taking a responsibility are (Harnish 2005, 38).

It should be noted that the kind of or responsibility or commitment considered here seems to just depend on the speaker's attitude. Even if it is a responsibility or commitment undertaken before an audience, nothing is said concerning the possible recognition or acceptance from the part of this audience. Harnish takes it that the responsibility (or commitment) has as its content "that certain conditions are satisfied" (p. 23), and this seems in conformity with the normative theories of Alston and Searle that he is discussing.

Nevertheless, there is an implicit step in Harnish argument that he does not discuss explicitly. Assuming that the conditions for its being the case that *P* are objective and make the speaker liable to assessment, there is still a categorical distance between the expression of a belief and the undertaking of a responsibility or commitment in performing this very act. Firstly, in many conversational contexts these normative stances seem to lie beyond the domain of epistemic responsibility (for example, committing the speaker to a coherent assumption of other assertions or beliefs not deductively related to the first one, or to a future course of action). In this sense, Harnish's analysis seems to be restricted to epistemic considerations and to a strict cognitive approach to the illocutionary act of assertion.

Moreover, the notion of being liable to criticism and blame conceptually entails deontic relations with other interactants, unless the notion of commitment (or responsibility) here considered is bound to what has been called "personal commitment" (Gilbert 2006, 133) or responsibility. A personal commitment is unilaterally created by a personal decision, intention or attempt of an agent, thereby he or she committing him or herself to it. And, to the extent that commitments can call for rescission, a personal one can be rescinded on one's own. Yet it seems very doubtful that the commitments or responsibilities undertaken by the speaker (and legitimately expected by the hearer) in the course of a communicative interaction can be restricted to this type.

3. Deontic affordances and joint commitment: the interactionalist approach

Many authors have considered this conception unsatisfactory, and this for different reasons. Originally, Austin had said that "the performance of an illocutionary act involves the securing of *uptake*" (1962,117). Clark (1996, 2006) has argued that this uptake is more than the mere understanding of a speaker's communicative intention. In his view, communication is a means to propose, negotiate, and carry out *joint projects*, that can be taken up (in the sense of recognized and accepted) by the addressee. For example, suppose Ann says "Please, sit down" to Ben, intentionally meaning that he is to sit down. According to Clark's analysis, her meaning is a type of intention that she cannot discharge without Ben doing his part in recognizing that intention: "For speakers to mean something, they must act jointly with their addressees" (1996, 153). Furthermore, Ben's acceptance allows us to see Ann's proposal as a joint construal by both of them.

Still, a further reason to regard Gricean accounts unsatisfactory is provided by the consideration that jointly committing to a joint project requires something more than epistemic and volitional mental states (like beliefs, common beliefs and intentions). Following Gilbert's notion of *joint commitment* (Gilbert 2000), Carassa and Colombetti have contended that "to create a joint

commitment it is necessary and sufficient that every party expresses his or her readiness to be so committed, in conditions of common knowledge" (Carassa and Colombetti 2009, 1842). Joint commitments are a particular class of social commitments, where social commitments are defined as desire-independent reasons for action that (contrary to other types of desire-independent reasons for action like moral or legal obligations) are intentionally created by the social interactants. In so far, they are viewed as networks of deontic relationships, intentionally created by the very agents that are connected by such relationships (p. 1843).

These notions allow for an interactional notion of meaning that takes into account the deontic stance of the interactants. While speaker's meaning can be understood as a function of personal communicative intentions, a new notion of *joint meaning* is defined in terms of the joint construal of speaker and hearer: "joint meaning is formed every time a speaker and a hearer jointly commit to the fact that a specific communicative act has been performed" (p. 1849). Moreover, it can be said that the main effect of a communicative act is to provide to the hearer with a *deontic affordance*, i.e., a perceived possibility of affecting the common network of deontic relationships that bind him or her to the speaker.

In the particular case of an assertion, this interactionalist account proposes that whenever a speaker asserts to a hearer that *p*, this act amounts to the speaker offering a propositional commitment to the hearer to the extent that *p* is true (where a propositional commitment is a commitment to a propositional content, pp. 1844, 1848). The difference with the intentionalist account becomes manifest when observing that a propositional commitment of this sort is a social commitment, although it does not presuppose, in general, a joint commitment: for the commitment undertaken by the speaker with the truth of *p* has not to be accepted by the hearer. The assertion can be seen as a *precommitment* offered by the speaker, i.e. a deontic affordance that the hearer can, or cannot accept.

Although the authors do not prosecute this analysis further, it can be noticed that conceptualizing the illocutionary act in terms of a deontic affordance provides a direct, nonproblematic connection between the level of the speaker's utterance and the interactional level in which he or she becomes liable to critical assessment. This normative connection was left insufficiently justified in Harnish intentionalist account. Furthermore, the strict epistemic approach to the type of commitments (or responsibilities) then in play is now seen to be extended to other dimensions of assessment, something also lacking in the Gricean approach. For these reasons, the interactionalist approach seems to provide with theoretical resources that are relevant for a satisfactory account of communicative interaction.

Notwithstanding this, it should be taken into account that this theoretical approach is not presented as a radical alternative to the intentionalist one. For it is accepted that there is speaker's meaning, entirely determined by his or her communicative intentions in Grice's sense, whenever he or she undertakes a precommitment by means of his or her utterance. Here, it can be said that the speaker has the reflexive intention to produce an effect on a hearer through the recognition of the intention itself. Yet the difference in the interactional approach lies on the consideration that the intended recognition should be qualified as a "shared recognition", that is, a common belief, as intended by the speaker (p. 1847). For only then the utterance can qualify as a precommitment, as a deontic affordance in need of a further action on the part of the hearer.

In spite of these arguments in favor of the interactionalist approach, however, it has difficulties that should be also recognized and confronted.

4. Some open questions and lines of research

A first difficulty present in Carassa and Colombetti's proposal is the categorical gap subsisting between the recognition and acceptance of a joint commitment, as formulated in purely deontic terms, and the cognitive processes through which speaker and hearer, individually, come to the conclusion that a specific communicative act has been performed, and are aware of their common conclusion. The problem is then to come to understand the elements of human mental architecture that underlie the human capacity to create joint commitments and desire-independent reasons for action in general. The interactional approach takes this capacity to go beyond the epistemic and volitional dimensions of cognition as these are usually embedded in available theories of human action. Carassa and Colombetti have recently suggested that besides the usual psychological types (beliefs and perceptions, desires and intentions), "it will be necessary to consider a further type of primitive psychological modes, of the normative type" (forthcoming, 11). This view seems compatible with the empirical theories put forward by Tomasello (2009) and other authors. This line of research, however, is in need of much development.

A second difficulty for the interactionalist approach concerns the need to provide a detailed analysis of every type of speech act, in terms of the network of deontic commitments (and possibly correlative rights or entitlements) that affect the interpersonal relationships of the interactants. Here, it should be taken into account that speech acts may have, in general, a potential of different illocutionary forces among which one is to be preferably selected by the interlocutors. It is not completely clear which criteria would allow them to select one among an array of potential deontic affordances, nor how could the theory discriminate between the illocutionary dimension and the

perlocutionary effect of a speech act, if Austin's original distinction is to be maintained. Concerning these aspects, much work is also needed before a comprehensive and sound theory of communicative interaction can be offered.

Finally, a third difficulty concerns the question whether an essentially deontic account of communication might be ultimately naturalized. A well balanced integration of conceptual analysis and empirical research, in the sense suggested above, might indicate the path to follow in this concern. There seems not to be a principled reason why a naturalistic approach to the peculiar normativity of communication and interpersonal action in general should not be possible. But the usual conceptual and theoretical devices, as available in most philosophical and cognitive theories and proposals, seem hardly adequate in this concern.

REFERENCES

Alston, W. 2000, *lllocutionary acts and sentence meaning*, Ithaca, Cornell University Press.
Austin, J. L. 1962, *How to Do Things with Words*, Cambridge, MA, Harvard University Press.
Bach, K. and Harnish, R. 1979, *Linguistic Communication and Speech Acts*, Cambridge, MA, MIT Press.
Carassa, A. and Colombetti, M. 2009, "Joint meaning", *Journal of Pragmatics* 4, 1837-1854.
Carassa A. and Colombetti, M. (forthcoming), "Creating interpersonal reality through conversational interactions", in Kobow, B., Schmid, B., and Schmitz, M. (eds.), *The background of social reality*, Berlin, Springer.
Clark, H. 1996, *Using Language*, Cambridge, Cambridge University Press.
Clark, H. 2006, "Social actions, social commitments", in Enfield, N.J. and Levinson, S.C. (eds.), *Roots of Human Sociality*, Oxford and New York, Berg.
Gilbert, M. 2000, *Sociality and responsibility: new essays in plural subject theory*, Lanham, Rowman & Littlefield, MD.
Gilbert, M. 2006, *A theory of political obligation*, Oxford, Clarendon Press.
Harnish, R. 2005, "Commitments and speech acts", *Philosophica* 75, 11-41.
Harnish, R. 2009, "Internalism and externalism in speech act theory", *Lodz papers in pragmatics* 5/1, *Special issue on speech actions*, 9-31.
Searle, J. 1969, *Speech acts: an essay in the philosophy of language*, Cambridge, Cambridge University Press.
Tomasello, M. 2009, *Why we cooperate*, Cambridge, MA, MIT Press.

EL ARTE DEL DIAGNÓSTICO: FALACIAS IRÓNICAS

ANTONIO DUARTE CALVO
UCM – Fundación Ramón Areces
antonovduartovich@hotmail.com

ABSTRACT: The paper will focus on intentional and contextual aspects which characterize discursive intervention and, therefore, should not make them vanish when addressing an aspect of discourse such as fallacies. I will try to clarify, by examples, the importance of context to judge any communicative act and I will address, finally, a curious type of fallacy (ironic fallacy) that serve as guidelines for establishing the complexity of certain fallacies, their ambiguity and their uncertainty if we do not respond to the multiple dimensions that characterize the use of arguments.

Fallacies are not autonomous textual objects, are the tip of the iceberg of something much more intricate and complex, if we do not consider the intentions and aspirations of those involved in the discursive context and the circumstances surrounding the course of the conversation, identifying fallacies can be extremely confusing and erroneous. Ultimately we could be victims of the ambiguity that can be released from a purely "logical'' analysis of discourse.

KEYWORDS: Fallacies, Philosophy of Language, Philosophy of Mind

Introducción

Centraré la ponencia en los aspectos intencionales (actitudes e intenciones de los que intervienen) y contextuales (marco del discurso y contexto de la intervención) que caracterizan toda intervención discursiva y que, en consecuencia, no deben escamotearse a la hora de abordar un aspecto discursivo como son las falacias. Trataré de esclarecer, mediante algunos ejemplos, la importancia que tiene el contexto para juzgar cualquier acto comunicativo y abordaré, por último, un tipo curioso de falacia (falacia irónica) que servirá de pauta para establecer la complejidad de determinadas falacias, su ambigüedad y su difícil delimitación si no se atiende a las múltiples dimensiones que caracterizan el uso de argumentos.

Las falacias no son objetos textuales autónomos, son la punta del iceberg de algo mucho más intrincado y complejo; si no consideramos las intenciones y pretensiones de los involucrados en el contexto discursivo, así como las circunstancias que enmarcan el curso de la conversación, la identificación de falacias puede ser extremadamente confusa y errónea. En definitiva, podríamos ser víctimas de la ambigüedad que se puede desprender de un análisis puramente "lógico" del discurso.

1. Importancia del contexto y de la intencionalidad

Veamos un ejemplo paradigmático de lo que trato de explicar. Imaginemos que Pedro se dirige a nosotros (auditorio) exclamando en tono quejumbroso: "¡Nadie me ha hecho daño! ¡Nadie me ha hecho daño!" Seguramente pensaríamos que Pedro ha perdido la razón o que, tal vez, sea un masoquista compulsivo que necesite sufrir indefinidamente. Pero si ampliamos nuestro campo de visión y hacemos un análisis de los antecedentes, nuestra impresión sobre Pedro será diametralmente opuesta. La explicación es sencilla: alguien cuyo nombre es Nadie ha agredido a nuestro antiguo masoquista[1].

Pero podríamos ir más lejos, porque resulta que nos hemos informado de que el violento Nadie no se llama en realidad así, sino que nuestro sofista ha utilizado esa estratagema para ganar unos segundos provocados por la confusión, y así poder huir. De modo que hemos pasado de considerar a Pedro como alguien que ha perdido la razón, a considerarle como víctima de una agresión, pero no loco; hasta nuestra nueva situación, en la que además de sufrir una agresión ha sido, y hemos sido, objeto de un engaño, precisando más, de un equívoco intencionado; y todas estas reinterpretaciones se han ido produciendo a la luz de nuevos descubrimientos en el contexto y en la intencionalidad de los actores.

Esta argucia (falacia de ambigüedad) no habría podido ser detectada sin un conocimiento preciso del contexto de comunicación.

Pero veamos algunos casos curiosos que, espero, terminen de enmarcar el asunto que nos ocupa, en donde las palabras se prestan a nuevos significados; nuestra interpretación dependerá de nuestro conocimiento de factores contextuales. En virtud de ese conocimiento adoptaremos, o no, una nueva hipótesis sobre su posible significado.

El primero de estos ejemplos, en relación con los malentendidos, hace referencia a una anécdota que se atribuye al escritor y poeta Francisco de Quevedo. Se cuenta, que a la reina Mariana de Austria, segunda esposa de Felipe IV de España, le enojaba mucho todo lo que hacía referencia a su cojera; parece que Quevedo se apostó el pago de una cena con algunos amigos, a que tenía el valor de llamarle coja en público. Su plan consistió en hacerse con dos ramos de flores: uno de rosas rojas y otro de claveles blancos. Una vez en la plaza pública, Quevedo se presentó a la reina, y tras realizar una reverencia, le ofreció los dos ramos de flores. A continuación, se dirigió a ella con estas galantes palabras: "Entre el clavel blanco y la rosa roja, su majestad escoja".

Con toda seguridad Quevedo cenó gratis esa noche.

[1] Se trata de uno de los pasajes de la *Odisea* reformulado; aquel en el que Ulises le dice a Polifemo que su nombre es Ninguno. Cuando Polifemo después de haber sido herido en su ojo pide auxilio a los otros cíclopes exclamando: "Ninguno me mata por dolo", éstos le responden que es imposible rehuir la dolencia.

Sus amigos sabían de la intención del propio Quevedo de insultar públicamente a la reina: por eso adoptarían una hipótesis alternativa a la ordinaria, es decir: "su majestad es coja". Los intérpretes, en este caso sus amigos, transforman el significado ordinario de esa afirmación con sentido: "escoger entre una rosa y un clavel" en una afirmación aparentemente sin sentido (obsérvese que una afirmación parecida a la emitida pero sin el juego de palabras podría ser: "Entre el clavel blanco y la rosa roja, su majestad es boba". Lo cual sería una afirmación a todas luces absurda), pero que en virtud del conocimiento del contexto, parece también adecuada a las intenciones de Quevedo.

Desde el contexto de la reina, el poeta le dio a escoger entre una rosa y un clavel, o tal vez no. Bueno, supongamos que la reina y el propio Quevedo no se llevaban bien. Supongamos que la reina ya había sido objeto anteriormente de sus afiladas mofas. En ese caso, y valorando esos contextos, la reina puede llegar a la conclusión de que está siendo insultada. De lo dicho se extraen dos ámbitos contextuales: 1. Por un lado el contexto del diálogo estará ceñido al protocolo relativo al trato de una reina y un súbdito en público; donde, por supuesto, la crítica, y mucho más el insulto, son faltas intolerables. 2. Por otro, el contexto de los participantes, en el que hay una evidente enemistad entre ambos y uno de ellos (Quevedo) es un maestro consumado en el uso de la ironía y los juegos de palabras.

Teniendo en cuenta estos contextos, lo que se plantea como un enigma o misterio es la interpretación coherente y ordinaria de la proposición. Es decir, lo inexplicable es la interpretación más inmediata, porque resultaría inapropiada dados los contextos. En concreto violaría nuestro segundo ámbito contextual. Me explico: para la reina no tendría sentido que el poeta quisiera regalarle un ramo de flores; su pensamiento tras ese ofrecimiento se podría traducir con una expresión como: "Aquí hay gato encerrado". Como afirma Davidson (en traducción de Marcela García):

> "el oyente se da cuenta que la interpretación estándar no puede ser la interpretación pretendida [...] lo absurdo o inapropiado de lo que el hablante habría querido decir si sus palabras se hubieran tomado de la manera 'estándar' alerta al oyente de que hay truco o algún error" (Davidson 1986, 434).

En cambio, la hipótesis del insulto, encaja con ambos contextos. Se puede camuflar como un ofrecimiento que satisfaga el contexto 1, y satisface también la lógica del contexto 2.

En un mundo en el que el esquematismo, el eslogan y la velocidad reinan sobre los discursos, hay que andar con mucho tiento en la interpretación discursiva para no ser víctimas de nuestra ignorancia (ignorancia discursiva); como aquella viuda de pastor protestante que, al día siguiente de morir su marido, recibió (por error de una letra en la dirección electrónica) el e-mail de

un hombre de negocios que se había citado con su mujer en un hotel de Miami y a quien había escrito con el *subjet* de: "Querida esposa: Ha sido un largo viaje, aunque vale la pena. Aquí ya tienen lista tu llegada para mañana. Pero prepárate, cariño, ¡hace un calor infernal!"[2].

2. Falacias irónicas:

Antes de acometer un ejemplo de lo que he llamado *falacias irónicas*, me gustaría hacer una distinción importante en lo que al uso de falacias se refiere. Es la distinción entre paralogismo y sofisma que tomo de Luis Vega (Vega, 2003). Paralogismo es un argumento erróneo pero con el sentido neutro de un fallo ocasional o un error cometido de buena fe; en cambio un sofisma es ya una estratagema o argucia argumentativa hecha a sabiendas con intención dolosa de probar algo, aunque a través de una prueba de suyo fallida.

Para introducir las *falacias irónicas* me ayudaré de un ejemplo expuesto por García-Carpintero en relación con lo que H. P. Grice llamó *implicaturas conversacionales* en su artículo *Logic and Conversation*.

El término 'implicatura' está relacionado con el de 'implicación' pero se diferencia de éste en que se refiere a expresiones que, a causa del contexto, dicen cosas que se alejan de su significado literal. El término 'conversacional' hace referencia a un intercambio entre al menos dos individuos que comparten ciertos intereses cooperativos.

En función de este objetivo Grice formula su principio cooperativo:

> "Haga usted su contribución a la conversación tal y como lo exige, en el estadio en que tenga lugar, el propósito, la dirección del intercambio que usted sostenga." (Grice 1975, 45).

Grice divide esta máxima en varias submáximas, a imagen de las categorías kantianas: máximas de cantidad, calidad, relación y modo.

Dentro de la categoría de cantidad (cantidad de información que debe ser dada) estaría la de hacer una contribución tan informativa como sea necesario. Dentro de las de calidad, "no decir lo que se crea falso" y "no decir aquello sobre lo que no se tiene los datos apropiados para decir que es verdadero". Dentro de las de modo "evitar la ambigüedad", "evitar la innecesaria prolijidad", "proceder de modo ordenado"...

Es evidente que muchas de nuestras conversaciones no están guiadas por estos principios, pero lo que es importante subrayar para nuestros intereses es que la violación de una de estas reglas nos ponga en alerta; es decir, que nos plantee un enigma.

[2] Blanco y Negro Cultural / 8-5-2004

El siguiente ejemplo, enunciado por García-Carpintero, aclarará lo dicho hasta ahora:

> "Begoña, quien tiene ciertas inclinaciones feministas, conduce el coche y yo la acompaño. Nos precede otro vehículo. El vehículo que nos precede realiza todo tipo de maniobras desafortunadas, de esas que exasperan a los otros conductores. Finalmente, Begoña encuentra la ocasión de adelantarle; al hacerlo, ambos miramos con curiosidad malsana al conductor del otro vehículo –quizás tratando de encontrar en su rostro signos inequívocos de incompetencia-. El conductor resulta ser una mujer. Entonces Begoña dice: '¡Vaya, un coche blanco tenía que ser!' ". (García-Carpintero 1996, 492-93).

La exclamación de Begoña plantea un enigma. Convencionalmente pone el acento en el hecho de que es un coche blanco el que ha realizado esas maniobras desafortunadas. Con la expresión 'tenía que ser', desliza una asociación entre este tipo de mala conducción y los coches de color blanco. Esta aseveración viola varias de las máximas conversacionales griceanas. Carece de los datos apropiados para hacer tal afirmación y no es pertinente, en el sentido de que es difícil establecer una asociación entre el color de un coche y la impericia del conductor. Por eso podemos hablar de que la aseveración de Begoña tiene toda la apariencia de la falacia de mala generalización, es decir, la inferencia desde un hecho comprobado hasta una generalización desmedida.

Supongamos que un tercer conductor hubiera visto lo ocurrido, oyendo las palabras de Begoña, pero desconociendo el sexo del conductor del vehículo blanco. Sin duda hubiera considerado extraña y falaz la exclamación de Begoña. ¿Cómo se hubiera podido explicar sus enigmáticas palabras, teniendo en cuenta los pocos datos contextuales que maneja? Podría haber conjeturado diversas hipótesis, por ejemplo: "la chica del coche de al lado habrá tenido un trauma con algún/algunos coche/s blanco/s; o tiene prejuicios supersticiosos con el color blanco, o...". Observemos la imposibilidad de confirmar las tentativas de explicación por parte de nuestro tercer conductor, con lo que quedaría en la incertidumbre, a causa de su escaso conocimiento pragmático.

Como vemos, se requiere, para resolver el enigma, un conocimiento más amplio, conocimiento que, si no se tiene, habrá que buscarlo en la medida en que sea posible; del mismo modo en que Sherlock Holmes encontraba cerillas hundidas en el barro para el pasmo de su querido Watson: "yo la encontré sólo porque la estaba buscando" exclamaba el insigne detective.

Por tanto, dados los elementos contextuales que yo manejo puedo proponer una explicación más plausible. Recordemos que, con respecto al tercer conductor, yo manejo dos datos claves:

1. El conductor del coche blanco es una mujer.
2. Begoña está en contra de ese tipo de prejuicios que asumen que las mujeres son inferiores a los hombres en ciertas capacidades, como la de conducir.

En situaciones como la que hemos descrito, muchos hombres hubieran dicho: "¡Una mujer tenía que ser!", para apoyar la hipótesis de que las mujeres conducen peor que los hombres. Por el segundo dato clave que yo manejo sé que Begoña no cree eso; y ella sabe, a su vez, que yo lo sé. Por eso adopto la hipótesis más que plausible de que en realidad, Begoña, hace uso de una falacia irónicamente (hay una relación causal entre conducir un coche blanco y conducir mal) para denunciar el uso de falacias de ese tipo (hay una relación causal entre ser mujer y conducir mal).

En definitiva, al denunciar la utilización de esa falacia me hace una denuncia implícita de todas las falacias de ese cuño (malas generalizaciones). En base a mi competencia abductiva, Begoña me está afirmando: "No hagas uso de generalizaciones espurias de ese tipo". El propósito suasorio de Begoña sería legítimo. Se podría hablar, por tanto, de que la supuesta falacia pronunciada en el contexto definido con anterioridad, es una falacia denunciatoria de todas las falacias con esa forma en contextos similares. Precisamente asume un contenido ridículo para alertar de una forma argumentativa espuria que es capaz de albergar una afirmación a todas luces ilógica (falaz). Si formalmente se puede hablar de falacia, pragmáticamente ya no.

Conclusión

El lenguaje en su uso es una fuente constante de desviaciones intencionales. No se puede recurrir a una taxonomía para comprobar lo que se dice, al igual que no se puede llevar a cabo un tratamiento estándar de las falacias. Identificar una falacia no supone necesariamente identificar un uso avieso (sofisma) o inconsciente (paralogismo) de la misma. Habrá que tener muy en cuenta, en ocasiones, el trasfondo emotivo y cognitivo compartido por el hablante y el oyente. La interpretación de estas desviaciones debe tener en cuenta muchos aspectos: los diferentes contextos (contexto de diálogo, contexto de los participantes), el uso convencional u ordinario del lenguaje, así como una capacidad abductiva, más o menos refinada según el caso, que se aproxima a la de los buenos investigadores.

Baroja relata en sus memorias que Unamuno decía que toda la práctica de la medicina estaba en aprender el *casillero*, con eso se refería al arte del diagnóstico. Afirmaba: "Tú ves, por ejemplo, una persona que tiene fiebre, dolor de costado, esputo rojizo... Se marcan en el *casillero* los tres síntomas y sale pulmonía, y se busca el remedio". Baroja calificaba este *casillero* de Unamuno como idea de portera, porque si fuera así, tan bien como los médicos sabios harían el diagnóstico las porteras, y explicaba:

> "Por poco que haya uno ejercido la Medicina, ha visto pulmonías sin esputo herrumbroso, sin fiebre alta y que matan al enfermo; fiebres tifoideas en que la elevación de la temperatura no tiene regularidad; gripes que acaban por arte de

> magia, cuando menos se lo espera uno; diabetes donde no hay hambre ni sed; dolores violentos en la región del apéndice que no son de apendicitis, etc., etc." (Baroja 1944, 268).

Confío en no caer en una falsa analogía si digo que algo así pasa con los usos del lenguaje que ya hemos caracterizado. No se puede recurrir a un muestrario sin tener en cuenta una serie de factores vivos que afectan y transforman el hecho comunicativo si no queremos convertirnos, en definitiva, en burdos investigadores.

REFERENCIAS

Baroja, P. 1944, *Desde la última vuelta del camino. (Memorias). Familia, infancia y juventud,* Caro Raggio, editor.

Davidson, D. 1986, *A Nice Derangement of Epitaphs. Philosophical Grounds of Rationality,* Editado por R. Grandy y R. Warner, Oxford University Press.

García-Carpintero, M. 1996, *Las Palabras, las Ideas y las Cosas,* Barcelona, Editorial Ariel, (2008)

Grice, H.P. 1975, 'Logic and conversation', in P. Cole & J. Morgan (eds) *Syntax and Semantics*, volume 3: Speech Acts, New York, Academic Press.

Vega Reñón, L., 2003, *Si de argumentar se trata,* Barcelona, Montesinos.

METAREPRESENTACIÓN Y DUALISMO EN LA LECTURA DE MENTES

VICTOR FERNÁNDEZ
Universidad de Granada

ABSTRACT: The mind-reading debate has been taken for two sets of theories: the simulationist theory and theory-theories. Both views have shared two assumptions: 1) a dualist distinction between cognition and behaviour, and 2) a metarepresentational structure of the mental states involved in mind-reading capacities. The aim of this paper is attempt to argue against those assumptions and propose an alternative view that I will define as an Ecological view of mind-reading.
PALABRAS CLAVE: lectura de mentes, *behaviour-reading, affordances*, dualismo comportamiento/cognición, metarepresentación, acción y percepción

1. La visión estándar de la lectura de mentes

Descartes concebía la mente como un entidad opaca, privada e inaccesible. Esta concepción le llevó a suponer una distinción metafísica entre cuerpo y mente que se conoce como dualismo cartesiano. La ciencia cognitiva asumió que este dualismo era erróneo, sin embargo mantuvo el supuesto de que la mente, al contrario que el comportamiento, era opaca, privada e inaccesible. El único modo de acceder a ella es mediante la observación del comportamiento, interpretación e inferencia. Esta presunción ha sido extrapolada al estudio de lectura de mentes. La lectura de mentes es la capacidad para predecir las acciones de un sujeto y su engloba habilidades como: la atribución de estados mentales, el modo de desenvolverse en contextos sociales por parte de un agente o el modo en que se manipula la mente de otros agentes para conseguir un objetivo dado. Siguiendo el dualismo entre comportamiento y cognición los teóricos de la lectura de mente han supuesto que los sujetos que tienen esta habilidad deben de acceder a la mente del otro, o bien mediante una teoría implícita sobre cómo funcionan los estados mentales de los demás que es contrastada con su comportamiento (theory-theory); o bien mediante un mecanismo que simula los estados mentales de los demás a través de los mecanismos propio-perceptivos del agente (simulation theory).

La visión también ha supuesto que los estados mentales involucrado en lectura de mentes tienen una estructura metarepresentacional. Dado que el acceso a la mente del otro es un proceso

por el cual se atribuye e interpreta el comportamiento del otro e infiere sus estados mentales, el producto de dicho proceso es un estado que representa la actitud del otro hacia sus propias representaciones. Los simulacionistas no presentan estas dos asunciones de un modo tan explícito. Sin embargo, la teoría simulacionista exige por un lado, la asunción de que el otro agente tienen estados mentales similares a los del lector; Y por otro, inferencias desde los estados producidos por la simulación off-line a la atribución de estados mentales al otro sujeto (Stone y Davies, 1996). Ambas asunciones parecen necesitar tanto de metarepresentación, como interpretación e inferencia para simular los estados mentales del otro y adecuarlos a su comportamiento.

Un ejemplificación de la incorporación de estos supuestos (dualismo y metarepresentación) puede encontrarse en Bermúdez (2003, 2009). Bermúdez hace dos distinciones en el modo de abordar la lectura de mentes: En primer lugar, distingue entre: 1) *minimal mind-reading* donde el sujeto coordina su comportamiento de manera sistemática con los estados mentales de los otros (El nombre más usado en la literatura es *behaviour-reading* y será el que use en adelante); 2) *substantive mind-reading*, donde el agente se representa los estados mentales del otro. En segundo lugar, dentro de estos *substantive mind-readings* podemos distinguir dos versiones según se tenga una capacidad metarepresentacional más o menos compleja: 1) *perceptual mind-reading*, se da cuando el agente se representa los estados perceptivos del otro. 2) *propositional attitude mind-reading* se da cuando el agente se representa las actitudes proposicionales del otro participante y está conceptualmente ligada al lenguaje. La asunción de que la mente es opaca e inaccesible viene dada por la distinción estricta entre animales que leen el comportamiento y animales que atribuyen mente. La asunción de que la lectura de mentes tiene una estructura metarepresentacional viene dada porque es necesaria para una *substancial mind-reading*, y porque la distinción entre *mind-reading* perceptual y proposicional está basada en la complejidad de dichas metarepresentaciones.

2. Una visión ecológica de la teoría de la mente

Existen algunas posiciones en Ciencia Cognitiva que de manera directa o indirecta ponen en duda ambos supuestos de la visión estándar. La primera es la corriente ecológica en psicología. Apoyada en las ideas de Gibson (1979), esta corriente ha trabajado bajos tres supuestos: 1) la percepción está ligada a la acción; 2) los eventos que percibimos contienen información que especifica las oportunidades para la acción del sujeto (*affordances*); 3) dichas *affordances* son propiedades de los objetos y eventos tomadas con referencia a un observador. McArthur y Baron (1983) han aplicado esta visión a la cognición social, de modo que un sujeto puede percibir a otro sujeto en términos de posibilidades de acción. El sujeto es capaz de extraer información útil para sí

mismo a través de reconocimiento de patrones de acción, movimientos corporales, miradas o gestos.

En segundo lugar, la perspectiva "amalgamada" (*amalgamated*) de la mente (Rowlands 2011) ha venido defendiendo dos tesis: 1) la mente es corpórea, depende constitutivamente del cuerpo; 2) la mente está embebida (embedded), los organismos explotan estructuras del entorno social y natural para llevar a cabo sus tareas cognitivas. Los estados mentales son representaciones con una modalidad específica dependiente de la relación entre el cerebro sensorio-motor y ciertas partes del cuerpo. Esta es la razón por la que debemos entender que percepción, pensamiento son constitutivamente interdependientes. Este enfoque implica que entender la mente como algo abstracto y despegado del cuerpo no tiene sentido y que por tanto, la mente no puede estar desligada de la acción y el comportamiento.

Por último, desde la psicología del desarrollo y la filosofía (Hobson 1993; Reddy, 2008; Gómez, 2004, 2005; Gomila 2001, 2002), han surgido distintas voces defendiendo lo que se ha conocido como la perspectiva en segunda persona de la atribución mental. Los defensores de esta visión se basan en una visión corpórea de la mente para defender que existe un modo de acceder a los estados mentales del otro que no tiene una naturaleza metarepresentacional. Este acceso se da de un modo conjunto entre dos sujetos (es recíproca) y está a la base de las relaciones de intersubjetividad. Dicho acceso en segunda persona se pone de manifiesto de una manera evidente en contextos comunicativos, relaciones cara a cara y episodios emocionales donde se involucran distintos sujetos.

Aunque estas tres líneas de pensamiento tienen su similitudes y diferencias, comparten el rechazo de los dos supuestos de la visión estándar. Además, de su corpus teórico se extraen cuatro supuestos básicos que definirán lo que voy a denominar una *visión ecológica* de la lectura de mentes: 1) La mente es algo constitutivamente ligado a la acción y al cuerpo. Un sujeto percibe en términos de *"affordances"* de modo que puede reconocer patrones "mentales" a través de las propiedades físicas de los objetos. 2) un sujeto puede reconocer la mente de otro a través de patrones de conducta, relaciones cara a cara, gestos o posturas corporales. 3) Dicha capacidad es perceptiva, y por tanto se caracteriza por tener una estructura no metarepresentacional. 4) Esta capacidad está ligada a la percepción y la acción, no están mediadas por interpretación, inferencia o simulación.

3. Un Problema metodológico de la visión estándar

La visión estándar ha llevado el supuesto de la opacidad de la mente al dominio de la cognición social sin justificación alguna. Esta visión no sólo ha no ha justificado el presupuesto que

de que la mente es inaccesible sino que ha supuesto de modo *ad hoc* que la opacidad de la mente es un problema para los organismos involucrados en contextos sociales. Este supuesto es *ad hoc* necesita de justificación tan pronto como nos damos cuenta de que los sujetos humanos leen la mente en contextos sociales de una forma automática e intuitiva. Es obvio además, que en contextos comunicativos, emocionales y relaciones cara a cara, los agentes no parecen estar prediciendo el comportamiento de otros o creándose representaciones de su mente (Nöe, 2009; Reddy, 2008). Por tanto, pensar que cuando un agente lee mentes está salvando una barrera infranqueable entre comportamiento y cognición es algo que no es evidente como *la visión estándar* pretende. Para defender que la lectura de necesita mecanismos de inferencia, metarepresentación o simulación, ante debe demostrarse que los organismos tienen el acceso restringido a la mente mediante percepción.

4. Problemas empíricos para la visión estándar

La visión estándar tiene además, problemas para acomodar ciertos estudios empíricos en psicología. McArthur y Baron (1983) exponen una batería experimentos a favor de una *visión ecológica* de la cognición social. Por un lado, los sujetos son sensibles tanto a estructuras variables como invariables de las propiedades físicas de otros agentes. Por otro lado, los observadores perciben dichas propiedades como pistas de rasgos de personalidad y actitud emocional que facilitan la acción (*affordances*). Archer and Arkert, (1977) demuestran que a través de la moción de agentes interactuando, un sujeto puede percibir relaciones sociales complejas. Además, Alley (1981) mostró cómo diferentes sujetos decían sentir mayor deseo de protección ante estímulos de personas con características físicas de bebé. Esto implica no sólo que los agentes pueden reconocer actitudes emocionales o rasgos de una persona a través de sus acciones, sino que dichas propiedades son *affordances* para el sujeto que las percibe. Los experimentos ponen de manifiesto la incoherencia de pensar en la mente como algo opaco e inaccesible, y la dificultad de hacer una distinción tajante entre behaviour-reading y substantial mind-reading. La visión ecológica implica que algunos estados mentales pueden percibirse como *affordances* sin necesidad de apelar a estructuras metarepresentacionales por parte del agente que las percibe. Además, estos casos no están recogidos por ninguna de las distinciones de Bermúdez. Por una lado, no pueden ser etiquetados como behaviour-reading, ya que implican algo más que reaccionar ante el comportamiento de un agente. Por otro lado, no existe una atribución de estados perceptivos en términos metarepresentacionales, sino que se trata de percibir patrones "mentales" a través de la conducta del otro.

Otro ejemplo problemático para la visión estándar es el de la atención conjunta. Esta capacidad de dos individuos de compartir su atención hacia un objeto aparece hacia los 9 meses de edad. La atención conjunta se divide en dos capacidades: "attention following" and "attention contact". La primera se basa en la capacidad del niño de percibir la atención de otro hacia un objeto. La segunda en la capacidad de percibir la atención del otro hacia uno mismo (Gómez, 2005). Estas capacidades parecen estar ligadas a la percepción de estados corporales específicos lo que implicaría que los sujetos no ven la mente como algo totalmente desligado del cuerpo. Además de la atención del otro, los niños pueden percibir las actitudes emocionales que este tiene hacia el objeto. Así lo explica Hobson (1993):

"Lo que observamos es que el Bebé percibe la actitud afectiva que se expresa a través de los gestos faciales, vocales, corporales en definitiva, de la otra persona y que reacciona a ella. Esta actitud, en cuanto que percibida, tiene al mismo tiempo un <<valor>> emocional y una dirección de sentido externo, con respecto a los sucesos, los objetos y las personas del mundo" (Hobson 1996, 62)

El bebé, no solo explota la información que hay en el comportamiento mismo, sino la información mental, emocional y expresiva percibida a través del comportamiento. Behne, Malinda y Tomasello (2005) demostraron que los niños eran capaces de entender la conducta del investigador como manifiestamente comunicativa (a través de miradas y gestos como apuntar). Para evitar una interpretación en términos de *behaviour-reading,* contrastaron estos casos con otros donde los gestos no se acompañaban con movimientos faciales o miradas directas a los ojos del niño. En estos casos el éxito de la tarea caía drásticamente. En otra variante de experimentos asociados a este (Liszkowski, Carpenter y Tomasello, 2007), se demostró que el niño es también sensible al interés y la actitud del adulto hacia el objeto, incluso si el objeto desaparece del campo de visión de ambos.

Existen varias conclusiones que podemos rescatar de los experimentos. En primer lugar, es poco plausible que el tipo de habilidades expuestas descansen sobre procesos psicológicos complejos, inferenciales, metarepresentacionales o simulacionales. En segundo lugar, la membrana que separa la mente y el cuerpo es más permeable de lo que la visión estándar supone. En tercer lugar, la captación de intenciones comunicativas por niños pre-lingüísticos es un caso que no caen dentro de las distinciones de Bermúdez. Por un lado, la intención de comunicar es un estado cognitivo (no puede ser reducido a *perceptual mind-reading*); por otro lado, no es un estado que podamos entender en términos proposicionales ni metarepresentacionales, como Bermúdez exige para casos de *propositional mind-reading*. Esto nos lleva a considerar la posibilidad de que, aunque

existan casos que caigan dentro de las distinciones de Bermúdez, es probable que la lectura de mentes sea un continuo con muchos más grados de los que ha supuesto la visión estándar.

5. Ventajas explicativas de la Visión Ecológica

La *visión ecológica* nos muestra la lectura de mentes como una cuestión de explotación de información física con respecto a un observador. Esto nos ayuda a explicar dos fenómenos difícilmente acomodables por *la visión estándar*. Por un lado, la gran variabilidad de estrategias usadas por los organismo en medios sociales. Existe una gran cantidad de especies animales que denominamos sociales. Estos animales exhiben estrategias, habilidades y organizaciones sociales muy distintas y variables. Un problema de la rigidez de *la visión estándar* es que presupone que las habilidades de estos organismos pueden reducirse a *perceptual mind-reading* o *behaviour mind-reading*. Es difícil ver como esta variabilidad es reducible a sólo dos capacidades. Sin embargo, tan pronto como tenemos una visión ecológica, dónde el organismo puede explotar diferente información del otro (conductual, mental, emocional, comunicativa), podemos promover diferentes explicaciones para diferentes nichos sociales sin presuponer ninguna habilidad de alto nivel, y atendiendo a rasgos importantes como el ambiente o las organización social.

En segundo lugar, hay un fenómeno que *la visión estándar* no ha tenido en consideración. Parece que hay una correlación entre la complejidad cognitiva de un organismo y la complejidad de estrategias que exhibe en su organización social. De este modo, animales socialmente complejos como los mamíferos superiores, parecen exhibir a su vez una mayor complejidad cognitiva. Este fenómenos es asumible por *la visión ecológica* en tanto que esta implica una co-evolución nicho-organismo. Cuanto mayor se la capacidad de percibir información mental por parte de un organismo, mayor complejidad exhibirá el nicho social, y viceversa. Por supuesto, los defensores de la theory-theory o la teoría simulacionista podrían esgrimir que aceptan dichos fenómenos, y que dicha correlación y variabilidad podría cristalizar en forma de mecanismos más o menos desarrollados. Sin embargo, los mecanismos de mind-reading, tal y como los entiende la visión estándar, descansan en estructuras metarepresentacionales que difícilmente pueden atribuirse a ciertas especies. Por otro lado, es más parsimonioso suponer que las habilidades sociales descansan en mecanismos perceptivos de reconocimiento de *affordances* que en mecanismos del alto nivel que necesitan de inferencia y meta-representación.

Conclusión

He intentado demostrar que la visión estándar de la lectura de mentes es problemática por

diversos motivos. En primer lugar, sus dos supuesto básicos tienen problemas tanto metodológicos, como para acomodar ciertas evidencias empíricas. En segundo lugar, la visión ecológica de la lectura de mentes se encuentra mejor posicionada, no sólo para evitar estos problemas, sino para dar cuenta de dos fenómenos biológicos: la variabilidad de estrategias sociales desarrolladas por diferentes organismos, y la correspondencia entre la complejidad cognitiva de un organismo y el grado de complejidad social del medio en el que esta inmerso. Aunque los defensores de la theory-theory o la visión simulacionistas tiene modos de resolver estos problemas apelando a grados de madurez de sus mecanismo, parece más parsimonioso apelar a mecanismo más básicos y que no necesite apelar a meta-representaciones, simulaciones o inferencias.

REFERENCIAS

Alley, T. R. 1981, *Caregiving and the perception of maturational status*, Unpublished doctoral dissertation, University of Connecticut

Archer, D. & Arkert, R. 1977, 'Words and Everything Else: Verbal and nonverbal cues in social interpretation', *Journal of Personality and Social Psychology* 35, 443-449"

Behne, T., Malinda, M. & Tomasello, M. 2005, 'One-year-olds comprehend the communicative intentions behind gestures in a hiding game', *Developmental Science 8,* 492 - 499

Bermúdez, J.L. 2003, *Thinking without Words*, Oxford, Oxford University Press.

Bermúdez, J.L. 2009, 'Mindreading in the animal kingdom' in Lurz, W.R. (Ed.) *The Philosophy of Animal minds*, MIT Press.

Carruthers, P. & Smith, K. P. (Eds.) 1996, *Theories of Theories of Mind*, Cambridge University Press

Gibson, J. J. 1979, *The Ecological Approach to Visual Perception,* Houghton Mifflin: Boston

Gómez, J.C. 2004, *Apes, Monkeys, Children, and the Growth of Mind*, Cambridge, MA: Harvard University Press

Gómez, J. C. 2005, "Joint Attention and Notion of Subjects: Insights From Apes, Normal Children, and Children with Autism". In Eilan, N., Hoerl, C., McCormack, T. & Johannes Roessler (eds.), *Joint Attention: Communication and Other Minds. Issues in Philosophy and Psychology*. Oxford University Press.

Gomila, A. 2001, 'La perspectiva de segunda persona: mecanismos mentales de la intersubjetividad', *Contrastes* 6: 65–86

Gomila, A. 2002, 'La perspectiva de segunda persona de la atribución mental', *Azafea* 1: 123-138

Hobson, P. 1993, *Autism and the Development of Mind*, Erlbaum.

Liszkowski, U., Carpenter, M., & Tomasello, M. 2007, 'Reference and attitude in infant pointing', *Journal of Child Language* 33, 1 – 20.

Nöe, A. 2009, *Out of heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness,* New York, Hill and Young.

Reddy, V. 2008, *How Infants Know Minds*, Cambridge, Harvard University Press.

Rowlands, M. 2010, *The New of Sciences of the Mind: From Extended Mind to Embodied Phenomenology*, MIT Press.

Stone, T. & Davies, M. 1996, 'The mental simulation debate: a progress report' in Carruthers, P. & Smith, K. P. (Eds.) 1996, *Theories of Theories of Mind*, Cambridge University Press, 119-40

NOMBRES PROPIOS Y TÉRMINOS DE GÉNERO NATURAL: UN EXAMEN DE DOS SUPUESTAS SIMILITUDES

LUIS FERNÁNDEZ MORENO
Universidad Complutense de Madrid
e-mail: luis.fernandez@filos.ucm.es

ABSTRACT: According to Kripke, two important similarities between natural kind terms and proper names are that both types of expressions are rigid designators and that they appear in identity statements that are necessary if true; Kripke denominates these types of statements containing natural kind terms "theoretical identities". Nevertheless, Kripke claims that the latter similarity follows from the former. In this paper I have a double aim. Firstly, I will contend that it can be held that natural kind terms are rigid designators their referents being universals. Secondly, I will argue that the conception of natural kind terms as rigid designators of universals hinders the establishment of the truth of theoretical identities and hence of their necessity.
KEYWORDS: rigidity, truth, necessity, identity, universals.

En las dos primeras conferencias de (1980) Kripke se ocupa fundamentalmente de los nombres propios, mientras que en la tercera, y última, atiende de manera especial a los términos de género natural, radicando uno de sus objetivos en alegar la existencia de ciertas similitudes entre estos términos y los nombres propios. Según Kripke, una de estas similitudes consiste en que tanto los términos de género natural como los nombres propios son designadores rígidos; otra radica en que ambos tipos de expresiones figuran en enunciados de identidad que, en caso de ser verdaderos, aunque son verdaderos *a posteriori*, son necesarios, es decir, necesariamente verdaderos, o sea, verdaderos con respecto a todos los mundos posibles. En el caso de los términos de género natural Kripke denomina a ese tipo de enunciados que son verdaderos *a posteriori*, pero necesarios, "identificaciones teóricas" e "identidades teóricas" – optaré por esta última denominación –, y entre los ejemplos de tales enunciados él menciona "Agua es H_2O" y "Oro es el elemento con número atómico 79". Ahora bien, Kripke sostiene que esa segunda similitud se sigue de la primera. Él afirma:

> "Las identidades teóricas, de acuerdo con la posición que sostengo, son […] identidades que involucran dos designadores rígidos y, *por tanto*, son ejemplos de lo necesario *a posteriori*." (Kripke 1980, 140; la primera cursiva es mía).

Mi objetivo en este escrito es doble. En primer lugar, alegaré que cabe mantener que los términos de género natural son designadores rígidos; a este respecto asumiré la concepción de que los referentes de esos términos son universales. En segundo lugar, aunque voy a conceder el

carácter *a posteriori* de las identidades teóricas, argüiré que la concepción de las identidades teóricas según la cual expresan la identidad entre universales obstaculiza el establecimiento de la verdad de las identidades teóricas y, por tanto, de su necesidad.

En la primera conferencia de (1980) Kripke introduce el término *designador* como denominación común de los nombres propios y de las descripciones definidas, los dos tipos de términos singulares que Kripke toma en consideración. La definición de designador rígido en (1980) se encuentra en las conferencias primera y segunda y, por tanto, aparece formulada *antes* de que Kripke se ocupe de los términos de género natural. Puesto que Kripke afirma que los términos de género natural son, como los nombres propios, designadores rígidos, pero no presenta *explícitamente* una definición de designador rígido para los términos de género natural, ha de asumirse que la definición de esa noción aplicable a este tipo de términos será una *extensión* de la propuesta para los términos singulares.

La definición de designador rígido contenida en la primera edición de (1980) – publicada en 1972 – y en su escrito (1971) puede dar lugar a distintas caracterizaciones. No obstante, en una respuesta a David Kaplan, Kripke afirma que la noción de designador rígido (para los términos singulares) pretendida por él es la siguiente:

> "[U]n designador *d* de un objeto *x* es *rígido* si designa *x* con respecto a todos los mundos posibles en los que *x* existe, y *no designa un objeto distinto de x con respecto a ningún mundo posible*." (Citado en Kaplan 1989, 569).

Esta definición de designador rígido deja abierta dos opciones. La primera es que un designador rígido designe el mismo objeto con respecto a todo mundo posible – independientemente de que el objeto exista en ese mundo o no –, mientras que de acuerdo con la segunda, designe el mismo objeto con respecto a todo mundo posible en el que el objeto exista, careciendo de referencia con respecto a cualquier otro mundo posible. Siguiendo una terminología hoy usual, propuesta inicialmente por Salmon en (1981), los designadores rígidos que satisfacen la primera caracterización son designadores *obstinados*, mientras que los que cumplen la segunda son designadores *persistentes*.

Aunque Kripke deja abierta esa doble posibilidad, la definición de designador rígido aplicable a los nombres propios pretendida por Kripke es la primera de ellas, pues en el Prefacio a (1980) afirma que considera los nombres propios como rígidos *de iure*. Un designador es rígido *de iure* si al fijar su referencia se *estipula* que su referente es el mismo independientemente de que estemos hablando del mundo actual o de un mundo posible distinto del actual; así los designadores rígidos *de iure* son, en la terminología de Salmon, designadores obstinados.

A pesar de que la mayoría de las descripciones definidas son *designadores no-rígidos*, Kripke admite que hay algunas descripciones definidas que son designadores rígidos, pero éstas no son rígidas

de iure sino rígidas *de facto*. En el caso de un designador rígido *de facto* no se estipula que hay un único objeto que constituye su referente con respecto a todo mundo posible, pero el predicado contenido en la descripción se aplica al mismo objeto "*en cada mundo posible*" (1980, 21, n. 21; mi cursiva). De acuerdo con esta caracterización, las descripciones definidas que sean designadores rígidos, es decir, designadores rígidos *de facto* serían igualmente designadores obstinados y, aunque la tesis de que todas las descripciones definidas rígidas sean designadores obstinados es cuestionable, en aras de la simplicidad la asumiremos en lo siguiente. No obstante, aunque los designadores rígidos *de facto* son un tipo de descripciones, voy a extender esta noción de rigidez a los *términos generales* con los que se forman tales descripciones, ya que, según Kripke, los términos de género natural son términos generales.

Una consecuencia de la rigidez de los nombres propios es que los enunciados de identidad *verdaderos* en los que el signo de identidad es flanqueado por dos nombres propios son *necesarios*, es decir, verdaderos con respecto a todo mundo posible. Un ejemplo famoso es el enunciado de identidad "Héspero es Fósforo". Este enunciado es verdadero, pues los nombres "Héspero" y "Fósforo" designan el mismo objeto, es decir, el planeta Venus. No obstante, puesto que los nombres propios son designadores obstinados, esos nombres designarán el planeta Venus con respecto a todo mundo posible, de lo que se sigue que el enunciado "Héspero es Fósforo" es necesario.

Llegados a este punto, la primera cuestión a plantear es si la definición de designador rígido para los términos de género natural y la concepción de Kripke acerca de sus referentes, es decir, de los géneros naturales permiten sostener que los términos de género natural son designadores rígidos.

A este respecto conviene retomar la definición de designador rígido para términos singulares mencionada y extenderla a los términos de género natural. La extensión más natural, y la *única* que tomaré en consideración, es la siguiente:

> Un designador *d* de un género natural *g* es rígido si designa *g* con respecto a todos los mundos posibles en los que *g* existe, y *no designa un género distinto de g con respecto a ningún mundo posible.*

Como ocurre con la definición de rigidez para los términos singulares, esta caracterización de rigidez para los términos de género natural deja abierta *dos* opciones, a saber, que designen el mismo género natural con respecto a todo mundo posible *o* que designen el mismo género natural sólo con respecto a todo mundo posible en el que el género exista, careciendo de referencia con respecto a todo otro mundo posible. No obstante, siguiendo la tesis de Kripke acerca de los nombres propios, adoptaremos también la primera alternativa con respecto a los términos de género natural, es decir,

asumiremos que los términos de género natural designan el mismo género con respecto a todo mundo posible, o sea, los consideraremos como designadores obstinados.

Por lo que concierne a la concepción de Kripke acerca del estatuto ontológico de los géneros naturales, conviene señalar que hay pasajes de (1980) que contienen ejemplos un tanto idealizados de cómo procederían las introducciones de los términos de género natural en los que se caracteriza la relación entre un género natural y las entidades del género como una relación de *instanciación* – véase (1980), 135-136 –, por lo que esas afirmaciones de Kripke llevan a suponer que él concibe un género natural como un tipo de universal, más aún, como un universal instanciado en entidades particulares, si bien *no* introduce apenas detalles al respecto.

No obstante, a la vista de nuestra caracterización de designador rígido para los términos de género natural cabría suscitar la cuestión de cuáles son las condiciones en las que un género natural existe en un mundo posible. Una posición basada parcialmente en algunas afirmaciones de Kripke y que adoptaré en lo siguiente es que un género natural existe en un mundo posible si y sólo si ese mundo contiene instancias del género,[1] de modo que las afirmaciones acerca de la existencia de un género natural en un mundo posible son *derivadas* de aquellas acerca de la existencia de instancias del género en ese mundo. En este sentido voy a adoptar dos *supuestos* adicionales. Por una parte, asumiré la siguiente *condición necesaria de identidad* de géneros naturales: si dos géneros naturales son idénticos, las instancias de cada género son las mismas en todos los mundos posibles; por lo tanto, para establecer que dos géneros naturales no son idénticos bastará establecer que hay algún mundo posible en el que las instancias de cada género no son las mismas. Por otra parte, puesto que las instancias de géneros naturales no son entidades que existen necesariamente, ya que podemos estipular mundos posibles que no contengan tales instancias, la *existencia* de dichos géneros tampoco será necesaria, sino *contingente*. Por tanto, voy a asumir que los géneros naturales no existen necesariamente.

Conviene señalar que la concepción de los géneros naturales como ciertos universales (sustancias, especies, etc.) permite sostener que los términos de género natural son designadores rígidos, de acuerdo con la definición de designador rígido para estos términos formulada anteriormente y con nuestro supuesto de que esos términos son designadores obstinados, pues dichos términos designarán con respecto a todo mundo posible los mismos universales (géneros naturales).

[1] Tal como Kripke se expresa en ocasiones (véase, por ejemplo, 1980, 136), él ha de sostener que el que un mundo contenga instancias de un género natural es al menos condición necesaria para que el género exista en ese mundo. Paso por alto consideraciones incluidas en la página mencionada acerca de si la aceptación de la existencia de un género natural puede depender de si ha sido identificado con anterioridad o posterioridad a otro, cuando la identificación de ambos se ha llevado a cabo en base al mismo tipo de entidades.

Ahora bien, como se ha señalado frecuentemente en la bibliografía sobre esta temática – véase, por ejemplo, Martí/Martínez-Fernández (2010) – la concepción de los referentes de los términos de género natural como universales instanciados en entidades particulares parece ser generalizable a los referentes de *todos* los términos generales que se aplican a tal tipo de entidades, lo que nos llevaría a admitir que no sólo los términos de género natural, sino también todos esos términos generales serían designadores rígidos. No obstante, llegados a este punto conviene introducir ciertas precisiones.

Como ya indicamos, Kripke afirma que los nombres propios, es decir, términos singulares semánticamente simples, son designadores rígidos, mientras que la mayoría de las descripciones definidas, es decir, términos singulares semánticamente compuestos, son designadores no-rígidos. De acuerdo con esto cabría afirmar que los términos generales semánticamente simples son designadores rígidos; así, por ejemplo, los términos "agua" y "oro" son designadores rígidos, pero también lo serán otros términos como "soltero" o "mesa". De este modo la rigidez no permite *distinguir* términos de género natural de otros tipos de términos generales. No obstante, puesto que considero que la anterior es la manera más natural de extender la rigidez de los nombres propios a los términos de género natural, aceptaré tal consecuencia.

Sin embargo, conviene puntualizar que *no todos* los términos generales serán designadores rígidos. Dicho de manera más precisa, la *mayoría* de los términos generales semánticamente compuestos no serán designadores rígidos. Así, por ejemplo, el término "metal más valioso" o "líquido preferido por José" no serán designadores rígidos – o, al menos, tienen una interpretación no-rígida –, pues pueden designar con respecto a distintos mundos posibles diferentes metales o líquidos, concebidos como universales. Esto no es óbice para que haya algunos términos generales semánticamente compuestos que sí sean designadores rígidos, como, por ejemplo, los términos "elemento con número atómico 79" o "sustancia compuesta de moléculas que constan de dos átomos de hidrógeno y un átomo de oxígeno", abreviadamente, el término "H_2O".

Admitido lo anterior, hemos de examinar si las identidades teóricas son enunciados necesarios. A este respecto tomaré en consideración el primer ejemplo mencionado de identidad teórica, el enunciado "Agua es H_2O", pues similares observaciones se aplicarían a la identidad teórica concerniente al oro y a otros ejemplos de identidades teóricas.

El enunciado "Agua es H_2O" es concebido por Kripke como un *enunciado de identidad* en el que el signo de identidad (de segundo orden) está flanqueado por los términos generales "agua" y "H_2O". Siguiendo la pauta del enunciado de identidad "Fósforo es Héspero", cabría alegar ahora igualmente que, puesto que los términos "agua" y "H_2O" son designadores rígidos – y más aún,

designadores obstinados –, si el enunciado de identidad "Agua es H_2O" es verdadero, será también *necesario*. No obstante, aquí se suscita un problema que no se planteaba con respecto al enunciado "Fósforo es Héspero", a saber, el problema de cómo podemos establecer la verdad y, por tanto, la necesidad del enunciado "Agua es H_2O". En el caso del enunciado "Fósforo es Héspero", y en general de los enunciados de identidad que contienen dos nombres propios, para establecer su verdad y, por tanto, su necesidad sólo se requiere tomar en consideración el mundo actual y averiguar si la entidad designada por ambos nombres en el mundo actual es la misma. Sin embargo, de acuerdo con nuestra condición necesaria de identidad de géneros naturales y, por tanto, de sustancias, que parece razonable, el establecimiento de la verdad de una identidad teórica como "Agua es H_2O" involucra la consideración de todos los mundos posibles.

Como indicamos anteriormente, hemos adoptado una concepción de los géneros naturales y, por ende, de las sustancias, según la cual las sustancias agua y H_2O son universales, por lo que el enunciado "Agua es H_2O" será *verdadero* si y sólo si las sustancias agua y H_2O son idénticas y, por tanto, de acuerdo con la condición necesaria de identidad de géneros naturales y, por tanto, de sustancias, esto será así *sólo* si las instancias de cada una de esas sustancias son las mismas en todos los mundos posibles, es decir, sólo si los términos "agua" y "H_2O" son coextensivos en todos los mundos posibles. Sin embargo, aunque concedamos que la extensión de estos dos términos es la misma en el mundo actual – lo que habría sido establecido *a posteriori* –, no podemos ir tomando en consideración todos y cada uno de los mundos posibles para determinar si los términos "agua" y "H_2O" son coextensivos en todos ellos.

Dicho de otra manera, hemos adoptado una concepción de los géneros naturales como universales, lo que permite que los términos de género natural sean designadores rígidos; de este modo hemos aceptado la rigidez de los términos "agua" y "H_2O" y, más aún, cabe conceder la coextensividad de esos términos en el mundo actual, pero de aquí no se sigue su coextensividad en todos los mundos posibles, que es condición necesaria para la verdad – y la necesidad – del enunciado "Agua es H_2O".

Esto puede ilustrarse de la siguiente manera. Tomemos en consideración una porción de agua que no existe en el mundo actual pero que existe en un *mundo no-actual.* De la rigidez de los términos "agua" y "H_2O", y de la coextensividad de estos términos en el mundo actual, *no* se sigue que dicha porción sea también una instancia de H_2O. De manera general, la rigidez de los términos "agua" y "H_2O" junto con su coextensividad en el mundo actual no son suficientes para establecer que las instancias de las sustancias designadas por estos términos sean las mismas en todos los mundos posibles y, por tanto, que el enunciado "Agua es H_2O" sea verdadero. Ahora bien, puesto que no

tenemos justificación para aceptar la *verdad* del enunciado "Agua es H_2O", tampoco tenemos razones para admitir su *necesidad*.

La conclusión a extraer es, por tanto, que, aunque la tesis de que los términos de género natural son designadores rígidos es sostenible, la concepción de los géneros naturales como universales, que posibilita que los términos de género natural sean designadores rígidos, obstaculiza el establecimiento de la *verdad* de las identidades teóricas y, por tanto, de su *necesidad*.

REFERENCIAS

Kaplan, D. 1989, 'Afterthoughts', en Almog, J. et al. (eds.), *Themes from Kaplan*, Oxford, Oxford University Press, 565-614.

Kripke, S. 1971, "Identity and necessity", en Munitz, M. (ed.), *Identity and Individuation*, Nueva York, New York University Press, 135-164.

Kripke, S. 1980, *Naming and Necessity*, Oxford, Blackwell. Reimp. revisada y con prefacio añadido de 'Naming and necessity', en Davidson, D. y G. Harman (eds.), *Semantics of Natural Language*, Dordrecht, Reidel, 1972.

Martí, G. y J. Martínez-Fernández 2010, 'General terms as designators: a defence of the view', en Beebee H. y N. Saabarton-Leary (eds.), *The Semantics and Metaphysics of Natural Kinds*, Nueva York, Routledge, 46-63.

Salmon, N. 1981, *Reference and Essence*, Princeton, Princeton University Press.

BOGHOSSIAN'S ARGUMENT FROM INFERENCE AND IDENTITY BELIEF ASCRIPTIONS

EKAIN GARMENDIA
UNAM

ABSTRACT: Boghossian proposed an argument against mental externalism, which claims that if externalism were true our *de dicto* ascriptions could not offer the rational explanations we expect from them. Here I present and defend an answer to this argument. I claim that the externalist can indeed offer good explanations if she makes use of identity belief ascriptions. In order to defend such a proposal from some possible problems, I superficially characterize a model about what it is to engage in an inference and what it is to offer a rational explanation.
KEYWORDS: Boghossian, Externalism, Inference, Rational Explanations.

1. The Argument from Inference

Externalism claims that content properties of our mental states are partly determined by factors external to us: what concepts a subject has, what she believes or thinks, depends on environmental factors. According to Boghossian's argument from inference, such a position is incompatible with our traditional model of rationality, and with our *de dicto* ascriptions offering rational explanations.

The argument is based on a slow switching example. It introduces Peter, who one day meets Pavarotti swimming in Lake Taupo. Peter is later transported to Twin Earth, which contains a *doppelgänger* of Pavarotti, Twin Pavarotti. Peter stays long enough on Twin Earth to acquire the concept TWIN PAVAROTTI, even though he never suspects there are two different individuals he dubs 'Pavarotti'. One day he attends a recital by Twin Pavarotti. The next day, Peter engages in a reasoning he would express by uttering the following sentences: 'I once met Pavarotti at Lake Taupo; I listened to Pavarotti last night; therefore, I once met at Lake Taupo the tenor I listened to last night.'

According to Boghossian, the externalist is committed to the following ascriptions, to saying that this is the inference Peter is engaging in:

(1) That Peter once met Pavarotti at Lake Taupo.
(2) That Peter listened to Twin Pavarotti last night.
(3) That Peter once met at Lake Taupo the tenor he listened to last night.

Boghossian points out that the example shows that externalism is incompatible with mental content being transparent. Transparency claims that we are always in a position to know whether two of our thoughts do share content or not. If the externalist's ascriptions are right, then the first premise in Peter's inference contains the concept PAVAROTTI, and the second premise the concept TWIN PAVAROTTI, while Peter believes that he expresses the same concept with all his utterances of 'Pavarotti'.

Now, according to Boghossian, if content is not transparent then the "*a priority* of our logical abilities" will be false. Peter's inference is invalid, while he wrongly takes it to be valid. Moreover, *a priori* reflection will not be sufficient for him to amend his error; what he needs to discover is that he is thinking about two different individuals, which he can learn only by empirical investigation.

On the other hand, if the *a priority* of our logical abilities is false, our *de dicto* ascriptions will not rationalize beliefs and behavior. We ascribe desires and beliefs, among other things, in order to explain why people believe and behave as they do. But the externalist's ascriptions seem not to rationalize Peter's belief. Believing that one met Pavarotti at Lake Taupo and believing that one listened to Twin Pavarotti last night is no reason to believe that one met at Lake Taupo the same person he listened to last night. Nonetheless, Peter is intuitively rational; he *ought* to believe that he once met at Lake Taupo the tenor he listened to last night. The externalist's ascriptions do not explain why Peter is acting rationally.

2. The Identity Premise Solution

The response I favor to this argument simply denies that the externalist lacks ascriptions that describe Peter as acting rationally—the only thing the externalist has to do is to add an identity belief to her ascriptions.

Peter's is a mistake of equivocation: he believes that there is one tenor where in fact there are two. And such scenarios are quite familiar. When one confuses two objects it is common to make use of identity belief ascriptions in order to offer rational explanations. The externalist then can explain why Peter believes that he once met at Lake Taupo the tenor he listened to last night by claiming that Peter believes that he once met Pavarotti at Lake Taupo, that Peter believes that he listened to Twin Pavarotti last night, and that Peter believes that Pavarotti is Twin Pavarotti. Hence, according to the response I favor, one can correctly describe Peter's inference in the following way:

(0) That Pavarotti is Twin Pavarotti.

(1) That Peter once met Pavarotti at Lake Taupo.

(2) That Peter listened to Twin Pavarotti last night.

(3) That, therefore, Peter once met at Lake Taupo the tenor he listened to last night.

The inference so described is valid, and therefore the corresponding ascriptions do indeed explain why Peter ought to believe as he does. The externalist's ascriptions yield good rational explanations.

An important feature of this response is its denying that transparency plays as important a role as Boghossian intended. *Even if* mental content happens not to be transparent, our *de dicto* ascriptions will offer rational explanations. And this is the case, I think, because even if content is not transparent it does not follow that something like the "*a priority* of our logical abilities" is false. Peter confuses his concepts PAVAROTTI and TWIN-PAVAROTTI, but it is far from clear that he will therefore engage in inferences that are invalid. So, it is far from clear that Peter's reasoning constitutes some kind of counterexample to the "*a priority* of our logical abilities".

3. First-Person Perspective and Inference Externalism

A first, quite common reaction to the response I favor is to protest that it fails to correctly describe Peter's first-person perspective: the explanation proposed above might quite correctly explain *why* Peter adopts his belief, *why* he is doing what he does, but it fails to correctly describe *what* Peter is doing. After all, Peter did not consciously consider any identity belief when engaging in his reasoning and, if he were to explain why he believes that he met at Lake Taupo the tenor he listened to last night, he would make use of no identity statement. The explanation I have proposed does not seem to correctly characterize Peter's perspective.

A second worry has to do with an alleged consequence of the solution. One might suspect that claiming that Peter's inference actually contains an identity premise would have as a consequence something we might dub 'Inference Externalism'.

Suppose that Schpeter is Peter's *doppelgänger*. Peter and Schpeter are internally identical, and they are disposed to utter exactly the same sentences. The difference between them resides in that Schpeter has never been the victim of any switch, he has never entered into contact with Twin Pavarotti. All of Schpeter's utterances of 'Pavarotti' refer to Pavarotti. And, as Peter, Schpeter engages in a reasoning to conclude that the tenor he once met at Lake Taupo is the tenor he listened to last night.

We would not use any identity belief ascription to explain Schpeter's belief. Now, if our

ascriptions are correct it seems that, even if they are internally identical, Schpeter is engaging in an inference that lacks any identity premise, while Peter is engaging in an inference that contains such an identity premise. That is: due to external factors, the inferences Peter and Schpeter are engaging in differ in their structural properties. One might think that this "Inference Externalism" is quite anti-intuitive, and it does not seem to be independently motivated.

I suspect that these two worries may be motivated by the acceptance of a faulty picture about what it is to engage in an inference (and what it is to offer a rational explanation), and that they disappear once we stick to a correct picture on these matters. The wrong model would assume that whenever we judge an argument sound we engage in an inference constituted by the very same premises and conclusion of the argument, and that offering a correct rational explanation consists in these cases in correctly identifying the premises and the conclusion of the argument considered. I will not discuss this model here. Instead, I shall make a couple of points on these topics in the following section, and shall later explain why the worries mentioned here are ungrounded.

4. Two Notes on Inference Engaging and Rational Explanation

The first point I would like to make is that engaging in an inference is for some of our beliefs to *cause* us to adopt a further belief. We have a bunch of beliefs, and because we have those beliefs, we many times feel ourselves committed to adopting some further beliefs. We can look at this in "causal terms": our holding to some beliefs causes us to adopt a further belief—and this constitutes *inferring* some belief from others.

A second point is that, when trying to offer a rational explanation of a belief one holds to, it is not our goal to identify any argument the subject of the explanation considered. Instead, our aim is to explain why the subject adopted the belief in question, identifying those beliefs that move her to acquire the new belief. Now, when we engage in an inference there is typically a whole net or heap of beliefs that are causally responsible for our adopting a further belief. Of course, when we describe someone as engaging in an inference, when we explain why she believes that p, we do not ascribe to her all the beliefs that played some causal role in her believing that p. We have to pick out, among the beliefs that caused her to believe that p, just those whose ascription is relevant for the explanation we are offering.

Suppose that the belief that p is causally relevant for S to adopt the belief that q. Then ascribing to S the belief that p in order to explain why she believes that q will be relevant if p and the other ascribed beliefs constitute a valid argument. The point is that we offer *rational* explanations of other's beliefs and behaviors, describe them as acting *rationally*, by making explicit

the logical and epistemic relations between their beliefs and the reasons that move them to adopt those beliefs.

Of course many times one tries to explain why one believes as one does. Sure enough, most of the times one is better positioned than others when it comes to explain why one believes that *p*. But there are exceptions. The relevance of an ascription for an explanation might depend on factors the ascriber is not in a position to access, and this is the case also when the ascriber is the very same subject ascriptions are made about. For example, the ascriber/ascribee might have wrong beliefs about the contribution of a proposition to the validity of the argument she is trying to identify.

This might happen, for example, if one makes wrong judgments about some semantic relations between propositions. As Peter's example suggests, having such wrong metasemantic beliefs might put one in a position of making wrong judgments about the logical form of a given argument and, with that, of making wrong judgments about its validity. If one makes wrong judgements about the form of a given argument, then one will perhaps fail when picking out, among the beliefs that played some causal role in adopting some given belief, those that are relevant for the explanation one is trying to offer.

An important point is that those offering a non-optimal explanation of their own beliefs are not necessarily behaving irrationally. Making wrong judgments about the validity of argumental forms and wrongly characterizing the logical form of an argument do not condemn one, by themselves, to irrationality. On the contrary, whether a subject is rational or not is determined by her inferential practices, by why she adopts the beliefs she does, not by which arguments she judges valid.

Let me now explain why the two worries I introduced above are ungrounded. I said that one might protest that the identity premise solution does not nicely show Peter's *first-person perspective*. I do not think that this worry is well motivated; the ascriptions I proposed in the second section do indeed offer a good rational explanation of Peter's belief. Concretely, his belief that Pavarotti is Twin Pavarotti plays such a causal role (if he did not believe that, he would not accept the conclusion of his inference), and the ascription of such a belief is relevant for explaining *why* Peter believes as he does: the validity of the argument the ascriber is forming depends on that belief constituting a premise. Sure enough, Peter would not describe the inference he is engaging in as containing an identity premise. But as I just said, one is not always in the best position to offer a rational explanation of one of her own beliefs. Since he ignores the fact that that there are two different concepts he expresses by the term 'Pavarotti', Peter fails to see that the ascription of such an identity belief is necessary for offering a correct rational explanation of his belief.

The identity premise solution offers a satisfactory rational explanation of Peter's belief. If one is asking for more than this when protesting that the explanation fails to describe Peter's first-person perspective, then I am afraid she is asking for too much.

On the other hand, if we adopt a correct model of inference engaging and rational explanation, we see that there is no real problem with Inference Externalism either. As I said, the goal of our ascriptions is not to identify an argument one is considering, but to explain why one adopts a belief in question. Since the relevance of an ascription might depend on factors "external" to the subject of the ascription, it may happen that two individuals are internally identical, that the inferences they engage in do not differ in their structural properties, but that, due to contextual factors, different ascriptions are needed in order to offer good rational explanations of their respective beliefs.

This is Peter and Schpeter's case. Peter makes a mistake of equivocation, Schpeter does not. And this suffices for an ascription of an identity belief to be relevant in one scenario, but not in the other. Now, it does not follow that there is any structural difference between the inferences they engage in. Actually, the inferences Peter and Schpeter engage in do not differ in their structural properties. Nothing like Inference Externalism seems to follow from the identity premise solution, then, once we adopt a correct model of rational explanation.

5. **A Risk of Regress?**

Let me consider a last worry. Some have claimed that the identity premise solution condemns us to an infinite regress. Let us accept, as the identity premise solution suggests, that the only thing Peter has to do in order to save his rationality is to add an identity premise to his inference. Of course, he would express this premise by uttering the sentence 'Pavarotti is Pavarotti'. But now it could be that the first token of 'Pavarotti' in the identity premise and the token of 'Pavarotti' in the first premise of the inference do not refer to the same object. To *make sure* that they indeed do so, according to the identity premise solution, he will have to add a further identity premise, and we have the very same problem again. The regress seems obvious.

However, this criticism misses the mark; it misunderstands what the identity premise solution says. The argument from inference does not claim that if externalism is true, then we are at risk of being irrational, and the identity premise solution does not respond that we can *ensure* our rationality if we add enough identity premises to our inferences. Instead, the argument from inference says that if externalism is true, then our ascriptions will not rationalize beliefs and behavior and, accordingly, the identity premise solution responds that the externalist does indeed

have at hand ascriptions of attitudes that do explain Peter's situation. And as far as such an explanation is correct, it is false that Peter cannot conform his thoughts and beliefs to the norms of rationality.

Peter has nothing to ensure; the challenge is not directed at him, but at the externalist that aims to explain why Peter adopts the beliefs and attitudes he does. According to the identity premise solution, we can explain Peter's situation by ascribing to him reliance on an identity premise, and there is no need of any further premise (or, therefore, any risk of regress) because there is nothing we (or Peter) have to *make sure*. Peter's concept PAVAROTTI *actually* refers to Pavarotti when it occurs in the identity premise and when it occurs in the first premise; there is no need to *ensure* this fact.

REFERENCES

Boghossian, P. 1992a, 'Externalism and Inference', *Philosophical Issues Vol. 2, Rationality in Epistemology.*

1994, 'The Transparency of Mental Content', *Philosophical Perspectives Vol. 8, Logic and Language*.

2011, 'The Transparency of Mental Content Revisited', *Philosophical Studies* 155.

Brown, J. 2004, *Anti-individualism and Knowledge*, Cambridge, MIT Press.

Burge, T. 1979, 'Individualism and the Mental', *Midwest Studies in Philosophy* 4.

1998, 'Memory and Self-Knowledge', in Ludlow and Martin (eds.): *Externalism and Self-Knowledge*, Cambridge, CSLI Publications.

Faria, P. 2009, 'Unsafe Reasoning: a Survey', *Dois Pontos* 5.

Mercier, H. & Sperber, D. 2011, 'Why do Humans Reason? Arguments for an Argumentative Theory', *Behavioral and Brain Sciences* 34.

Pérez Otero, M. 2010, 'El debate entre externistas e internistas acerca de la racionalidad inferencial', *Análisis Filosófico* 30.

Schiffer, S. 1992, 'Boghossian on Externalism and Inference', *Philosophical Issues Vol 2, Rationality in Epistemology.*

Sorensen, R. A. 1998, 'Logical Luck', *The Philosophical Quarterly* 48.

Stalnaker, R. 2008, *Our Knowledge of the Internal World*, Oxford, Oxford University Press.

A TALE OF TWO MARY'S

HEIMIR GEIRSSON
Department of Philosophy
Iowa State University
Ames, IA 50011
U.S.A.

ABSTRACT: Anti-physicalists have in recent years presented interesting thought experiments when advocating the view that conscious states are something over and above the physical, or the view that the mental cannot be fully explained in purely physical terms.[1] To support their view they have largely relied on two thought experiments, one of zombies, which are human replicas that lack consciousness or phenomenal feel, and another of a color deprived scientist, Mary, who after a lifetime in a black and white environment is exposed to colors and sees a red tomato. The anti-physicalists have argued that the two thought experiments work best in tandem. The thought experiment about Mary and the argument based on that thought experiment, they argue, is best understood as easing the way for the zombie thought experiment and the argument based on it. However, upon closer inspection the two thought experiments seem to undermine and not support each other.

This paper will look at a tension between the two thought experiments. One way to remove the tension is to focus on the role of conceivability in the zombie thought experiment, suggesting that we are not *justified* in believing that zombies are conceivable.[2]
KEYWORDS: zombies, physicalism, conceivability

1. Zombies

David Chalmers initially presented the zombies as being physically and psychologically identical to us. However, none of their functions will be accompanied by conscious experiences. There is nothing it is like being a zombie.[3]

Why does Chalmers say that zombies are conceivable? It is obvious to most of us that a mile-high unicycle is logically possible. Chalmers writes "I confess that the logical possibility of zombies seems equally obvious to me." (Chalmers 1996, 96) So, Chalmers claim is that because the

[1] See especially David Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, 1996)., Frank Jackson, "Epiphenomenal Qualia," *Philosophical Quarterly* 32(1982)., and "What Mary Didn't Know," *Journal of Philosophy* 83(1986). Granted, not every physicalist might be affected by Chalmers' argument.

[2] I develop an account of conceivability and the role of justification more fully in Heimir Geirsson, "Conceivability and Defeasible Modal Justification," *Philosophical Studies* XLIII(2005).

[3] Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*: 95.

descriptions of the conceived scenario appear to be coherent the scenario conceived is logically possible, i.e., it is true in some possible world.

In its simplest form Chalmers' argument against materialism based on the zombie thought experiment, the *modal argument*, is as follows:

1.Zombies are conceivable.

2.If something is conceivable then it is possible.[4]

3.So, zombies are possible.

4.If zombies are possible, then materialism is false.

5.So, materialism is false.

Premises (1) and (2) need some support, and some support for (1) comes in the guise of Mary and the knowledge argument.

2. Mary

Frank Jackson's thought experiment centers on Mary, a super-scientist who is locked in a black and white room and who has never experienced other colors. Because her research focuses on how various subjects experience colors she knows everything there is to know about the physical processes of such experiences. She has, for example, discovered which wavelength combinations stimulate the retina when one observes the sky, how they produce via the central nervous system the workings of the vocal cords and the expulsion of air that results in the uttering of the sentence "the sky is blue." Suppose now that Mary is released from her black and white room and sees colors for the first time, e.g. a red tomato. Will she learn anything new? Jackson claims that it is obvious that she learns something new about the world and our visual experience of it. Even though she has all the physical information about what goes on in a person's body when she sees a red tomato, Mary will learn something about the *subjective* aspect of seeing red when she first sees a red object. She learns *what it is like* to see red.

The thought experiment and the intuition it generates can be captured in the following argument, *the knowledge argument*.

1.Mary has all the physical information concerning human color vision before ever experiencing a color.

[4] A weaker premise would state that conceivability is a good guide to possibility. Chalmers explores that issue in later works, and other, for example Steven Yablo, have done the same. See Stephen Yablo, "Is Conceivability a Guide to Possibility," *Philosophy and Phenomenological Research* Liii, no. 1 (1993).

2.When experiencing a color Mary obtains knowledge that she did not have before, namely knowledge about the subjective aspect of experiencing the color.
3.So, not all information concerning human color vision is physical information.
The argument generalizes about other experiences, that is, in general we have the intuition that the subjective aspect of experience is something that is not captured with purely physical information.

3. Zombies and Mary

So what are the relative roles of the zombie thought experiment and the Mary thought experiment? One way to understand the arguments is that each one works independently of the other and that each one is sufficient to undermine logical supervenience.[5] But this understanding is challenged in a passage by Chalmers when the focus turns from logical supervenience to the falsity of materialism.

> Taking the knowledge argument alone: most materialists find it hard to deny that Mary gains knowledge about the world, but often deny the step from there to the failure of materialism. ... In perhaps the most powerful combination of the two arguments, we can use the knowledge argument to compellingly establish the failure of logical supervenience, and the modal argument to compellingly make the step from that failure to the falsity of materialism. (Chalmers 1996, 145-146)

And in "The Two-Dimensional Argument Against Materialism" he states that it is obvious that the knowledge argument, as presented by Jackson, cannot defeat physicalism.[6] However, the modal argument can do so, according to Chalmers.

4. Zombie Mary

The physicalists introduce Zombie Mary in an attempt to show that the modal argument and the knowledge argument create a tension and so undermine each other instead of being mutually

[5] In order for B-properties to logically supervene on A-properties, someone who knows all A-facts will be able to ascertain B-facts about the situation from the A-facts alone . Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*: 70.

[6] David Chalmers, "The Two-Dimentional Argument Against Materialism," in *Oxford Handbook of the Philosophy of Mind*, ed. Brian McLaughlin (Oxford University Press, 2009). Chalmers points out that the term 'physical facts' in Jacksons argument is ambiguous and that while Mary know all the narrowly physical facts, namely all the facts about microphysics, she might not know the broadly physical facts that supervene on the narrowly physical facts. Materialism only requires that phenomenal facts are broadly physical facts.

supportive. The tension between the two arguments is brought out in the following way.[7] Imagine Zombie Mary, who is a physical duplicate of Mary. Like Mary, upon first leaving the room, Zombie Mary claims that she has had a new experience and she reports having learned something new, namely something about what it is like to see red. But because she is a zombie she cannot be right about this. Being a zombie it cannot be the case that she has learned anything new about the subjective aspect of experiences when seeing the red tomato for the first time because the relevant subjective aspect is missing from the life of zombies. So there is a clear difference between Mary and Zombie Mary. Mary claims to have learned something new about her subjective experiences and her claim is true, while Zombie Mary's claim to the same effect is false due to the fact that she lacks the relevant experiences. All the information there is to have about Zombie Mary is physical information.

In Mary's case the anti-physicalist has the intuition that Mary learns something new about the subjective aspect of experience when she experiences color for the first time. In Zombie Mary's case, the anti-physicalist has to accept that Zombie Mary seeming to learn something new about subjective experiences when she first experiences color does not show that she has learned something new about such experiences. Instead, a purely physicalistic account captures all there is to know about Zombie Mary and her experiences.

The anti-physicalist is now in an unfortunate position. She has to admit that in certain situations the modal argument allows one to disregard or ignore the intuition elicited by the Mary thought experiment. The intuition elicited by the Mary thought experiment is that Mary learns something new when she first experiences colors, and what she learns has to do with the subjective aspect of experiencing colors. But when Zombie Mary exits the black and white room and enters the world of colors she learns nothing new in spite of her claims to do so. The reason is, of course, that she does not have subjective experiences.

The tandem of the knowledge argument and the modal argument was supposed to work in such a way that the knowledge argument paved the way for an acceptance of a key premise in the modal argument, namely that zombies are conceivable. But once we acknowledge that zombies are conceivable, then that very possibility undermines the intuition that powers the knowledge argument. We can no longer trust the intuition that tells us that one learns something new about the subjective experiences when experiencing color for the first time.

[7] The tension between the two arguments is presented and discussed in Victoria McGeer, "The Trouble with Mary," *Pacific Philosophical Quarterly* 84 (2003).

Given the tension between the modal argument and the knowledge argument it seems evident that the anti-physicalist has to sacrifice one of the two. The question is, which one should go? Of the two argument it is the modal argument that is needed to conclude that materialism is false, and the modal argument, while helped along by the knowledge argument, does not require the knowledge argument in order to succeed. Further, it is the introduction of zombies that generates the problem for the intuition that drives the knowledge argument. It is therefore reasonable to investigate the modal argument and the thought experiment that drives it. I believe that a closer look at the zombie thought experiment reveals that we are not justified in believing that zombies are conceivable and so the modal argument fails.

5. The conceivability of zombies

A crucial premise in the modal argument claims that zombies are conceivable. I question the conceivability claim. My take on the conceivability claim is significantly weaker than that of other critics. Instead of denying that zombies are conceivable, i.e., instead of claiming that one encounters contradictions when trying to conceive of zombies, I shall argue that one is not *justified* in believing that zombies are conceivable. The latter is significantly weaker than the former, that is, one might not be justified in believing a conceivability claim even though there are no contradictions to be found in the conceived scenario. Nevertheless, if one wants to question the modal argument, then it is sufficient to argue that one is not justified in believing one of its premises.

When Chalmers introduces zombies he claims that it seems obvious that they are possible, just as it seems obvious that a mile high unicycle is possible. If the conceivability claim is questioned he says that there is little to offer as a reply other than repeating the description of the conceived scenario and asserting its obvious coherence. As Chalmers writes, "[T]he burden is on the opponent to give us some idea of where the contradiction might lie in the apparently quite coherent description."[8] However, there are mistakes here. First, there is significantly more to coherence than the absence of contradictions and so a conceivability claim can be undermined without one finding a contraction in the conceived scenario. Second, there are good reasons to claim that in many, perhaps most cases, it is the one who makes the conceivability claim who should provide evidence for it being true when questioned about it.

One significant difference between the mile-high unicycle example and the zombie example is that the former involves a common object that is enlarged, a function that everyone is familiar

[8] Chalmers, *The Conscious Mind: In Search of a Fundamental Theory*: 99.

with, while the latter involves a very complex scenario that is to large extent described with the help of technical terms. The mere complexity of zombies indicates that there is quite a bit of work to be done to show that once can coherently conceive of them. In the absense of showing so, a key premise of the modal argument is without support.

Consider Zombie Mary again as she leaves the room and sees the red tomato, thus experiencing a color other than black and white for the first time. Being neuron-for-neuron identical to Mary, Zombie Mary will claim that now she knows what it is like to see red and now she knows what she was missing when she was working on the neurophysiology of color perception. The problem is that these claims are false, Zombie Mary being a zombie, and Zombie Mary does not know that they are false. Things become worse when Zombie Mary jams her finger when trying to reach for the tomato. Being neuron-for-neuron identical to Mary, Zombie Mary cries out as if in pain and claims that her finger hurts. Again, what she says is false. But here we have a further problem as it is not even clear that the words 'pain' and 'hurts' have any meaning in the mouth of Zombie Mary. Certainly 'pain' does not refer to pain, the phenomenal feel, in her mouth as there is no such pain to refer to. Perhaps we could claim that 'pain' refers to a brain state or a brain function. That does not seem very feasible. First, it is implausible to hold that 'pain' refers to a brain state or function in Mary's mouth (or our non-zombie mouths). Second, we run into difficulty when we try to account for Zombie Mary's talk when she says that her pain is becoming unbearable It seems that something becomes unbearable only because of how it feels, or because of what it is like to have the relevant experience. Why would a state or a function become unbearable if there is nothing it feels like to be aware of it? Finally, even though zombies do not have phenomenal states, the anti-physicalist wants to grant that zombies have intentional states and so can direct their talk appropriately. She then needs to explain how it is that it appears possible for the zombies to behave as if these terms have meanings. Given that our normal criteria for applying phenomenal feel terms can no longer be based on behavior or even neurological events it is unclear what kind of an explanation can be provided for the meaning and use of our phenomenal feel terms.

We are encountering here a certain lack of coherence that Chalmers calls "delightful and disturbing" but "not obviously fatal to the non-reductive position." (Chalmers 1996, 181) Chalmers' claim is understandable given that his focus is on avoiding contradiction. But while we might not at this point be able to derive a contradiction from the more complete description of zombies, there should be nothing delightful about the lack of coherence of the anti-physicalists' point of view. The decrease in coherence introduced to the zombie world when inquiring into the meaning of phenomenal terms and the truth value of Zombie Mary's honest claims that she had learned

something about her subjective states when seeing the red tomato, as well as her honest claims about her pain, is great enough for one to question whether zombies can be coherently described. While we have not uncovered a contradiction in the description of zombies, we have raised enough concerns to question whether we are *justified* in believing that they are conceivable.[9]

REFERENCES

Chalmers, D. 2009, 'The Two-Dimensional Argument against Materialism', in McLaughlin, B. (ed.), *Oxford Handbook of the Philosophy of Mind*, Oxford University Press.
Chalmers, D. 1996, *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.
Geirsson, H. 2005, 'Conceivability and Defeasible Modal Justification', *Philosophical Studies* XLIII, 279-304.
Jackson, F. 1982 'Epiphenomenal Qualia', *Philosophical Quarterly* 32, 127-36.
Jackson, F. 1998 'Postscript on Qualia', in *Mind, Method and Conditionals: Selected Essays*. 76-79. London: Routledge.
Jackson, F. 1986 'What Mary Didn't Know', *Journal of Philosophy* 83, 291-95.
McGeer, V. 2003 'The Trouble with Mary', *Pacific Philosophical Quarterly* 84, 384-93.
Yablo, S. 1993 'Is Conceivability a Guide to Possibility', *Philosophy and Phenomenological Research* Liii, 1-42.
[10]

[9] It is interesting that Jackson, the original proponent of the knowledge argument, has questioned the argument and the intuition that seems to support it. See Frank Jackson, "Postscript on Qualia," in *Mind, Method and Conditionals: Selected Essays* (London: Routledge, 1998).

[10]

ON QUOTATION AND SEMANTIC STRUCTURE

MARIO GÓMEZ-TORRENTE
Instituto de Investigaciones Filosóficas, UNAM, México DF 04510, Mexico
mariogt@unam.mx

ABSTRACT: Davidson observed that quotations must somehow get interpreted in a general way that exploits the salient pre-referential relation between a quotation and the expression between its quotes. On Herman Cappelen's and Ernie Lepore's "minimal theory" of quotation, the Davidsonian observation requires quotations to be semantically structured. In this communication I argue briefly that the Davidsonian arguments of Cappelen & Lepore have no implication that quotations must be semantically structured.
KEYWORDS: Quotation, Semantic structure, Reference-fixing, Use-mention

Davidson observed that quotations, unlike typical names, must somehow get interpreted in a general way that exploits the salient pre-referential relation between a quotation and the expression between its quotes, rather than via individual naming stipulations that need not exploit any such relation. On Herman Cappelen's and Ernie Lepore's (2007) "minimal theory" of quotation, the Davidsonian observation requires quotations to be semantically structured; together with some widely accepted objections to early descriptivist/demonstrative theories (see e.g. my objections in Gómez-Torrente 2001, or Cappelen & Lepore's similar objections in 2007, 117ff.), this leads them to the idea that quotations can be neither semantically unstructured terms nor descriptions, and must instead be strange structured but non-quantificational noun phrases, whose exact compositional structure Cappelen & Lepore leave undescribed.

In this brief note I want to argue that the Davidsonian observation, as well as some related new observations by Cappelen & Lepore (2007, pp. 101-3), have no implication that quotations must be semantically structured. I will indicate, especially by appeal to examples of classes of terms in natural language, that there is no inconsistency in supposing that the terms in a certain class are semantically unstructured and yet get their referents fixed wholesale with the help of a general rule that exploits a pre-referential relation between terms and intended referents. Thus those observations pose no obstacle to the view that quotations form a class of terms of precisely this kind.

It may be useful to remember first that semantic unstructuredness and wholesale reference-fixing are widely thought to form a consistent combination, indeed an actualized one. On a widely accepted view, natural language has some unstructured context-sensitive terms which nevertheless get their contextual referents fixed via general conventional rules. Pure indexicals like "I" are widely believed to be examples of this combination. (See e.g. the considerations in Kaplan 1977, 520.) Still, if we restrict our attention to pure indexicals, one might legitimately complain that wholesale reference-fixing has not been shown to be consistently combinable with unstructuredness *in the absence of context-sensitivity*, while the objections to descriptivist/demonstrative theories mentioned above and proposed both by Cappelen & Lepore and by me seem to imply that no context-sensitivity is involved in the case of quotation. But I think that, once one bears clearly in mind the distinction between fixing the reference of a term and giving its semantic structure, the idea that context-insensitivity might preclude the unstructuredness of a class of terms whose referents are fixed wholesale should have little appeal. Provided that a certain rule has been given that plays the essential role in fixing the referents of the terms in a certain class, what is to prevent the reference fixer from additionally convening that the terms are to be used as unstructured, and in particular as not equivalent with any description extractable from the rule? Nothing, it would seem. And it would also seem that one can create artificial examples without difficulty. In fact, Kaplan's "Dthat(α)" terms (cf. e.g. Kaplan 1977, 521-2) form an artificial class of terms which are context-insensitive (at least if instances of "α" are restricted to context-insensitive descriptions) and unstructured by explicit convention. One can easily create other examples.

This shows, I think, that wholesale reference-fixing is consistently combinable with unstructuredness in the absence of context-sensitivity. It's just a short way to making it equally plausible that semantic unstructuredness and wholesale reference-fixing *exploiting pre-referential term-referent relations* are consistently combinable. I think it's clear that even some classes of terms from natural language show by example that there is no inconsistency in the thought. I will consider them at the same time as Cappelen & Lepore's arguments for inconsistency.

Cappelen & Lepore's "observation 2" is that a quotation cannot be unstructured because it is non-conventionally "proximate" to its quoted expression, as it stands in a non-conventional relation to it that can be used to get information about it. In my view, this is arguably irrelevant to the matter of semantic structure. Think of an old Spanish name like "Miguel López Rodríguez", introduced under the general convention that a name N_1 N_2 N_3 is to name the individual born on the feast day of Saint N_1 to the N_2-N_3 family. Like a quotation, "Miguel López Rodríguez" is also in a non-conventional relation to its referent, namely, the non-conventional relation *x is a name where the*

forename is that of a saint whose feast day is the day when y is born. (Of course it is conventional that a saint gets a certain name and a certain feast day, but these conventions are previous to and independent from the fact that a certain person is born on that day, so the fact that a name and a person stand in that relation is non-conventional.) And in virtue of this "Miguel López Rodríguez" can provide information concerning the person talked about by means of it, information that will be obvious to any person who knows that it is an old Spanish name and what the convention for fixing the reference of old Spanish names is (namely, the information that the referent is the person born on the feast day of Saint Michael to the López-Rodríguez family). But evidently this doesn't mean that "Miguel López Rodríguez" is semantically structured. A numerical street name like "32" is also non-conventionally related to its referent, namely, by the non-conventional relation *x stands for a number that gives the place in the ordering of the streets for the street y*. And in virtue of this "32" can provide information concerning the street talked about by means of it, information that will be available to users of numerical street names who are familiar with numerical street names rules (namely, the information that the referent is, or at some point was or was thought to be, the 32nd street in the ordering of the parallel streets that began at a certain point in a certain direction). But numerical street names are plausibly unstructured devices. Similarly, it could be argued, a quotation is certainly non-conventionally related to its referent, and can provide information concerning its referent to speakers apprised of the relevant convention. But this is compatible with the quotation itself being semantically unstructured.

Cappelen & Lepore's "observation 3" is that if a quotation had no semantic structure, there would be no rule for interpreting it. Again the comparison with the cases of old Spanish names, numerical street names, and others, indicates how this is a non sequitur. There exist appropriate rules that serve to interpret those terms, but this is compatible with the terms being unstructured, because they arguably work only at the reference-fixing level, not at the level of semantic structure. Similarly, in quotation the relevant rule may well provide an interpreting canon that works merely at the reference-fixing level, leaving semantic structure as an independent matter.

Cappelen & Lepore's "observation 1" is perhaps their most novel objection to the unstructuredness thesis. It's the claim that the relationship between a quotation and the thing *x* that it quotes cannot be the relationship between a semantically unstructured expression and what it refers to, because any semantically unstructured expression could have been used to refer to *x*, but no quotation other than the one quoting *x* could have been used to quote *x*. A brief reflection completely dissolves this objection, however. Quite generally, if a class of terms get their referents fixed with the help of a general rule, and in such a way that two different terms in the class never

get the same referent, then it will be true that if *t* is a term of the class that refers to an object *o*, then no term in the class other than *t can* refer to *o*, given that *t* is the term determined as the name of *o* by the rule, and that the rule has been conventionally adopted. The "can" is of course a deontic modal, as in Cappelen & Lepore's objection. For example, no old Spanish name other than "Miguel López Rodríguez" could have been used to refer to Miguel López Rodríguez, given that he was born on the feast day of Saint Michael to the López-Rodríguez family and that the general rule for introducing old Spanish names had been instituted. The case of quotations may well be similar: it is quite true that no quotation other than ' "Socrates" ' could have been used to refer to "Socrates", but this is merely because once the reference-fixing rule for quotation has been adopted, the reference of a quotation other than ' "Socrates" ' will not be "Socrates". Thus no implication that the intuition of deontic modality is grounded in the structuredness of quotations is warranted.

REFERENCES

Cappelen, H. and E. Lepore 2007, *Language Turned On Itself: The Semantics and Pragmatics of Metalinguistic Discourse*, New York, Oxford University Press.

Gómez-Torrente, M. 2001, "Quotation Revisited", *Philosophical Studies* 102, 123-153.

Kaplan, D. 1977, "Demonstratives", in J. Almog, H. Wettstein and J. Perry (eds.), *Themes from Kaplan*, New York, Oxford University Press, 1989, 481-563.

UNA NOCIÓN DINÁMICA DE LA REPRESENTACIÓN MENTAL EN EL MARCO DE UN REALISMO NATURAL

EDUARDO M. GONZÁLEZ DE LUNA
Universidad Autónoma de Querétaro
(Mexico)

ABSTRACT: Perception is a process that involves representations, of which the perceiver is not always aware. From a natural realism, it is proposed that direct perception of the objects of the physical world is consistent with a model of mind that involves representations. It is suggested a deflationary notion of mental representations as operations, activities, modes or states produced in the perceptual process. As predecessors of this dynamic notion of representation, it is presented the concepts of sign in Thomas Reid and Charles Peirce.
PALABRAS CLAVE: representación mental, percepción, realismo natural.

La concepción cartesiana clásica de las mente asume que las representaciones mentales son los objetos de la percepción inmediata, constituyendo interfaces entre nuestros poderes cognitivos y el mundo exterior, en lo que Hilary Putnam ha llamado la "concepción interfaz" de la mente dentro del marco de un "realismo metafísico".

Esta concepción interfaz ha originado problemas sobre la relación entre la mente y el mundo, como por ejemplo, el del "genio maligno" cartesiano, que en términos filosóficos contemporáneos se formula como el problema de los "cerebros en cubetas", el cual consiste en la imposibilidad de responder a la pregunta de si somos o no cerebros metidos en cubetas conectados a una computadora, la cual transmite en nosotros la información con la cual construimos nuestra realidad. Como ha mostrado Putnam (1981), esta pregunta sólo podría en principio plantearse si se presupone un cierto tipo de realismo metafísico, es decir, una concepción metafísica del mundo real, que lo conciba como separado de nuestras capacidades cognitivas. Pero la respuesta satisfactoria del problema sólo podría darse "desde fuera", es decir, si lográsemos salir de la cubeta y adoptar una perspectiva externa, como la del "Ojo de Dios", lo cual sería, en principio, imposible desde nuestra condición de cerebros en cubetas, que nos obliga a permanecer confinados en la cubeta, con una perspectiva interna o mental. En consecuencia, el problema de los cerebros en cubetas es

irresoluble para tales seres, por lo que deberían considerarlo un problema originado en un planteamiento erróneo.

Ésta y otras paradojas escépticas que surgen del realismo metafísico muestran que desde esta postura se llega irremediablemente a la construcción de modelos en los cuales la relación entre las estructuras del mundo y las de las entidades mentales (*sense data*, *qualia*, apariencias u otras representaciones), aun cuando sea causal, no es epistémica o cognitiva, es decir, estos modelos nos dejan sin una justificación satisfactoria para explicar la relación de correspondencia entre la estructura del mundo y la de la mente.

La metafísica realista afirma que nuestros conceptos refieren porque corresponden con las estructuras del mundo independiente. Pero cómo se da esta correspondencia resulta ser un misterio. El misterio consiste en que una vez iniciados los procesos cognitivos, no hay nada que garantice la correspondencia entre la estructura de los esquemas conceptuales que son productos de nuestros procesos cognitivos y la estructura de la realidad, aunque sea ésta misma la causa que pone en movimiento los procesos cognitivos. Para que se establezca la correspondencia, dice Putnam, tenemos que suponer que hay ligaduras misteriosas o "rayos noéticos" que conecten ambas estructuras.

Putnam localiza el problema escéptico de la relación mente-mundo en la "concepción interfaz" de la mente. Por esto, propone una alternativa, el regreso al realismo natural (Putnam 1999), que es un tipo de realismo ingenuo en el cual tenemos acceso directo a la realidad mediante la percepción, la concepción y las otras capacidades cognitivas, pero sin tener que prescindir de la noción de representación, ni de la tesis de que la experiencia involucra la intervención de representaciones. Desde luego que la noción de representación que podría mantenerse en el marco de un realismo perceptivo directo como es el realismo natural difiere de la noción tradicional de representación como interfaz, pues debería ver a la representación no como un objeto de la percepción, sino como parte de un proceso cognitivo, una operación o actividad mental, o un estado intermedio particular de la mente considerada ésta como espacio representacional.

El propósito de este trabajo es señalar posibles caminos para formular una noción de representación que sea compatible con el realismo natural. Cabe decir que históricamente ha habido defensores de esta postura, e incluso en tiempos recientes, han surgido nuevas interpretaciones (Yolton 1975, 1984, 1987) que consideran que el propio Descartes, fue un defensor del realismo directo.

Como un antecedente de dos propuestas en esta dirección presentaré brevemente algunas ideas sobre los signos, las sensaciones y las percepciones, en la teoría de la percepción del filósofo

del siglo XVIII Thomas Reid para después referirme a la teoría de los signos de Charles S. Peirce, propuestas que, si bien son muy diferentes entre sí, a mi juicio sugieren nociones de representación adecuadas al espíritu del realismo natural, ya que consideran a las representaciones como parte dinámica de los procesos cognitivos que nos permiten tener acceso a la realidad, sin hacer una reificación de las mismas.

La teoría de la percepción de Thomas Reid (1785) distingue entre sensaciones, percepciones y otros tipos de representaciones, como recuerdos, juicios, etc. Reid concibe estas representaciones, no como interfaces de entidades presentes ante la mente, sino como operaciones, procesos, actos o estados mentales. En vez de reificar a las representaciones, mantiene un realismo directo con respecto a la percepción y los demás productos de los procesos representacionales mencionados. La mayoría de las representaciones tienen un contenido intencional, es decir, llevan a la mente a poner su atención en algún objeto distinto de la representación. Por ejemplo, en la percepción, o en la memoria, hay un objeto de atención distinto de la operación en sí misma: no podemos ver, sin ver algo; ni recordar, sin recordar algo. Pero la operación escapa a nuestra consideración, lo cual quiere decir que somos conscientes del objeto pero no del proceso de su producción, por lo que ambos deben ser cosas distintas.

Sin embargo, una de las características de la sensación es que esta operación no tiene objeto distinto de sí misma, a diferencia de la percepción y otras operaciones mentales. Para Reid la sensación es un acto de la mente que carece de un objeto diferente del acto mismo: Cuando un dolor se siente, el dolor sentido y mi sensación de dolor son lo mismo. Cuando un dolor no se siente no tiene existencia, y esto puede aplicarse a las demás sensaciones como colores, olores, etc.

La confusión de los filósofos cartesianos que los lleva a reificar las sensaciones a costa del objeto externo surge de que, en condiciones normales, la percepción y la sensación están unidas y no es fácil separarlas. Las sensaciones son parte de los procesos representacionales, son estados intermedios o momentos particulares de los mismos, pero no por eso debemos verlas como los objetos de aprehensión en la percepción del mundo externo. Para Reid, las sensaciones tienen el papel de signos de los objetos externos, dice que "Las sensaciones de olor, sabor, sonido y color, son infinitamente más importantes como signos o indicaciones que lo que son por propia cuenta; como las palabras de un lenguaje, en donde no atendemos al sonido sino al sentido." (Reid 1764, 25)

Esta analogía aclara el papel de las sensaciones en el proceso de la percepción. Como los sonidos de las palabras de un lenguaje hablado, no percibimos a las sensaciones sino directamente a

los objetos externos. Si bien la existencia de las sensaciones como signos, es parte fundamental de los procesos perceptivos, no son las sensaciones los objetos directos de la percepción.

El modelo reidiano de la percepción apoya el argumento que he presentado más arriba: aunque en la percepción estén involucradas representaciones mentales, como operaciones o procesos, o estados de los mismos, el objeto de la percepción es el objeto externo, y no la sensación, o algún otro tipo de representación, que sea signo de aquel objeto. Con este modelo se puede defender un realismo directo desde una teoría representacional de la mente.

Es importante notar que el argumento de Reid no busca negar la existencia de las representaciones mentales como estados internos, sino que su crítica se dirige a la pretensión de reificar a las representaciones a costa de negar la existencia de las cualidades de los objetos externos. Para Reid, esta tendencia surge del significado dual de las expresiones que usamos para hablar de las sensaciones. Pues usamos la misma palabra, por ejemplo "rojo", para referirnos a la sensación como una experiencia subjetiva y a la sensación como cualidad del objeto que se percibe mediante la sensación.

Este doble uso de los términos de sensación explica la aparente infalibilidad en la aprehensión de las sensaciones. Cuando las captamos, no es posible que las sensaciones posean características que en realidad no tengan. No puede ser que un hombre sienta un dolor que no tenga y no sea real. Reid pone énfasis sobre el compromiso realista acerca de la las sensaciones cuando son captadas en la experiencia introspectiva, en la cual las sensaciones no forman parte de los objetos del mundo externo. La aprehensión de la sensación involucra un compromiso con la existencia interna de la sensación; esto permite a Reid ligar a la sensación con la experiencia introspectiva; mientras que la percepción es experiencia del mundo externo.

La epistemología de Reid influyo en la filosofía primera de Charles Peirce en dos temas fundamentales, a saber, 1) en su semiótica o teoría de los signos desde una perspectiva realista, que afirma la inmediatez de nuestro conocimiento del mundo exterior, desde una cierta perspectiva, y 2) en su teoría de la fijación de las creencias y la dinámica del conocimiento. La teoría de Peirce reconcilia dos tesis aparentemente opuestas: afirmar que percibimos directamente cualidades de las cosas individuales y afirmar que la percepción y el conocimiento de los objetos que nos rodean es producto de procesos representacionales.

La influencia de la teoría de las sensaciones como signos de los objetos externos de Reid, se manifiesta en el énfasis de Peirce en el proceso representacional que niega la reificación de los signos. Es decir, ni Reid ni Peirce adoptan una concepción interfaz del signo y la representación. En particular en Peirce, su tesis de que las creencias son concepciones de disposiciones para actuar

y no meras representaciones mentales de la realidad es claramente un rechazo a la concepción interfaz cuya consecuencia natural es un giro hacia el realismo directo con respecto a la captación de ciertas cualidades de los objetos individuales.

Peirce (1877) afirma que de los varios métodos que existen para investigar, sólo el método de la ciencia está basado en la experiencia y presupone la existencia de la realidad, es decir, de objetos reales, independientes de nuestras opiniones, que afectan a nuestros sentidos según leyes regulares. El método científico supone que podemos saber cómo son esos objetos, y que cualquier hombre con la suficiente experiencia y razón llegará a la misma conclusión por lo que posibilita el acuerdo entre todos los hombres. Por tanto, la noción de investigación científica de Peirce que sirve para investigar cualquier aspecto de la realidad, tiene como su fundamento un realismo en el que el intelecto conoce directamente las cualidades de los objetos individuales por medio de una intuición inmediata que es confusa en un primer momento. Pero que posteriormente capta por abstracción hacia lo universal lo que es ese objeto expresado en términos de relaciones, tanto entre los estados del propio objeto, como con otros objetos, es decir, mediante su comportamiento.

Con respecto a la semiótica, al igual que Reid, la noción central para Peirce no es la de signo, sino la del proceso representacional mediante el cual algo funciona como signo y que supone tres elementos: signo o representamen, objeto y significado o interpretante. Peirce define el signo como "algo que está por algo para algo". Así pues, la semiosis se trata de una relación ternaria entre el signo o representación, el objeto representado y el interpretante de la representación.

Para Peirce la realidad sólo puede conocerse a través del signo porque está semiotizada. Todo lo que existe es signo, en cuanto que tiene la capacidad de representar y ser representado, de mediar y llevar ante la mente un significado. La semiosis, es el proceso triádico por el que algo puede funcionar como signo, esto es, tiene necesariamente que hacer referencia a un objeto y determinar un significado.

Además, Peirce (1897a, 99-100) distingue tres categorías de dimensión fenomenológica que llama primeridad, segundidad y terceridad. La primeridad se refiere a lo inmediato, a lo posible. Es el modo de ser de aquello que es tal como es, positivamente y sin referencia a otra cosa. La segundidad se refiere a las múltiples acciones y reacciones, a lo actual; es el modo de ser de aquello que es tal como es, con respecto a una segunda cosa. La terceridad se refiere a la mediación; es el modo de ser de aquél que es tal como es, al relacionar una primera cosa y una segunda. Es la categoría propia de la representación, el significado y el pensamiento. Las nociones de signo, objeto y significado se corresponden con las categorías de primeridad, segundidad y terceridad, respectivamente.

Estas categorías están presentes en todo fenómeno, en particular se distinguen en su máxima pragmática: "Consideremos qué efectos, que puedan tener concebiblemente repercusiones prácticas, concebimos que tiene el objeto de nuestra concepción. Entonces, nuestra concepción de esos efectos es el todo de nuestra concepción del objeto" (Peirce 1878, 38) Es así que, en el conocimiento del signo intervienen la máxima pragmática y las tres categorías fenomenológicas.

Según Rivas (1996), el dinamismo del modelo de Peirce se muestra cuando las categorías de signo, objeto y significado se definen únicamente por su posición lógica en el momento del análisis, esto es, de ser un primero, pueden pasar a ser un segundo o un tercero en otro momento de análisis. Y así puede darse un intercambio de posiciones lógicas dando lugar a una "semiosis ilimitada", en la que todo signo interpreta a otro signo y todo pensamiento da origen a otro pensamiento, Con respecto a los objetos del signo, Peirce distingue el objeto dinámico y el objeto inmediato. El objeto dinámico, u objeto mediato, es el objeto exterior al signo, es la realidad a la que el signo se refiere y que determina el signo a su representación. El objeto inmediato es el objeto interior al signo, el objeto tal y como es representado por el signo y cuyo ser depende de su representación en el signo.

Una noción central para dar cuenta del papel de la realidad exterior al signo en el proceso semiótico es la del "fundamento" del signo: El signo o *representamen* es "algo que está para alguien en lugar de algo bajo algún aspecto o capacidad. Se dirige a alguien, esto es, crea en la mente de esa persona un signo equivalente o quizá un signo más desarrollado. Ese signo creado es al que llamo interpretante del primer signo. Este signo está en lugar de algo, su objeto. Está en lugar de algo no en todos sus aspectos, sino sólo en relación con alguna idea a la que a veces he llamado la base (ground) del representamen" (Peirce 1897b, CP. 228).

Rivas (1996) interpreta la noción del fundamento como aquello que permite que el signo nos proporcione conocimiento acerca del mundo, desde una cierta perspectiva. Pues el signo representa a su objeto no en todos sus aspectos, sino por referencia a una idea, que es el fundamento del signo. El fundamento es un atributo del objeto dinámico en la medida en que dicho objeto se ha elegido de determinada manera y solo algunos de sus atributos se han elegido como pertinentes para la construcción del objeto inmediato del signo. Es debido al fundamento, que el signo se refiere a su objeto en algún aspecto, representando la realidad de un modo parcial y perspectual. Pero es la realidad extralingüística, exterior al signo, lo que determina al signo a que la represente de una determinada manera, no en su totalidad, sino siempre parcialmente, y sólo de esta manera se puede acceder a ella.

REFERENCIAS

Putnam, H. 1981 *Reason, Truth and History* Cambridge, Cambridge University Press

Putnam, H. 1999, *The Threefold Cord. Mind, Body and World* Columbia, University Press. [Versión en castellano de Alvarez Alvarez J. *La trenza de tres cabos. La mente, el cuerpo y el mundo* Madrid, Siglo XXI, 2001]

Reid, T. 1764 *An inquiry into the Human Mind on the Principles of Common Sense,* en http://www.earlymoderntexts.com/rein.html [Última consulta, el 20 de mayo de 2012]

_____ 1785 *Essays on the Intelectual Powers of Man,* en http://www.earlymoderntexts.com/reip.html [Última consulta, el 20 de mayo de 2012]

Buchler, J. 1955 *Philosophical Writings of Peirce* New York, Dover Pub.

Peirce, Ch. S. 1877 "The Fixation of Belief", en Buchler, J. (ed.) *Philosophical Writings of Peirce* New York, Dover Pub., 5-22

Peirce, Ch. S. 1878 "How to Make our Ideas Clear ", en Buchler, J. (ed.) *Philosophical Writings of Peirce* New York, Dover Pub., 23-41

Peirce, Ch. S. 1897a "What is a Sign? Three Divisions of Logic", en Buchler, J. (ed.) *Philosophical Writings of Peirce* New York, Dover Pub., 98-101

Peirce, Ch. S. 1897b "On Signs" [traducción al castellano de Mariluz Restrepo] "Fundamento, objeto e interpretante", http://www.unav.es/gep/FundamentoObjetoInterpretante.html. [Última consulta, el 20 de mayo de 2012]

Rivas M. U. 1996 "Frege y Peirce: en torno al signo y su fundamento", *Anuario Filosófico*, (29), No. 3, 1211-1224 http://hdl.handle.net/10171/533 [Última consulta, el 20 de mayo de 2012]

Yolton, John W. 1975, "Ideas and knowledge in the Seventeenth-Century Philosophy" in *Journal of the History of Philosophy*, vol. XII, 2, 318-330

Yolton, John W. 1984, *Perceptual Acquaintance from Descartes to Reid* Minnesota, University of Minnesota Press.

Yolton, John W. 1987, "Representation and Realism: Some Reflections on the Way of Ideas" in *Mind*, 96 (383), 318-330

INFERENTIALISM, METAPHOR AND CONTEXT

JAVIER GONZÁLEZ DE PRADO SALAS
UNED
jgonzalez@bec.uned.es

ABSTRACT: The purpose of this paper is to propose an inferentialist account of metaphor talk. In order to do so, it will be suggested that the inferential relations governing the use of expressions may be context sensitive. In particular, the substitutional inferences that determine the semantic role of sub-sentential expressions would behave in a different way in metaphorical and literal uses of the expression.
KEY WORDS: inferentialism, metaphor, contextualism, compositionality, normativity.

Metaphor has usually been considered a heterodox linguistic phenomenon. The difficulties of traditional theories with metaphor are sometimes related to their deficiencies in accounting for the role of context in linguistic practices. In this paper, I intend to study how inferentialist semantics can deal with metaphor and with the related contextualist worries.

According to the dominant tradition in philosophy of language, the meaning of a sentence is compositional: it is determined systematically by the meaning of its components and the way they are combined. This explains several key features of language use, in particular its productivity (the capacity of understanding an indefinite number of new sentences). Compositionality is usually associated with a bottom-up view of semantics: content is assigned firstly to sub-sentential expressions, and from there we will construct the meaning of whole sentences.

Metaphor poses a direct challenge to this traditional program; for in metaphoric expressions, there seems to be a divergence between how we actually understand the sentence and the literal meaning we would obtain computing its compositional content. Different answers to this challenge have been suggested. Many authors have tried to find a solution compatible with standard compositionality. For instance, from the field of Gricean pragmatics, it is proposed that, in metaphoric talk, the speaker says something in order to communicate something else: the compositional linguistic content expressed by the utterance differs from the speaker´s meaning. This approach is typically combined with a two-step model of metaphor processing. The hearer firstly interprets the metaphorical utterance literally; if this literal reading fails to fulfil some conversational principle, the hearer will try to find an alternative intended meaning for the utterance. When this model is presented as a psychological account of metaphor interpretations, it

has to face experimental results showing that, in many occasions, the processing time of metaphorical discourse is not longer than that of corresponding literal utterances.

Davidson, on the other hand, has argued that there is no special metaphorical meaning – not even as something merely communicated by the speaker. The only content associated with metaphors is their literal compositional meaning, usually false or trivially true. What is characteristic of metaphors is not their content, but how they are used to make us see one thing as something else: this use cannot be captured by any propositional meaning. However, this position fails to account for the assertive role of many metaphorical utterances in everyday talk.

The problems arising in the elaboration of accounts of metaphor consistent with standard compositionality have been exploited by contextualist authors in their criticism of traditional semantics. Contextualism argues that contextual factors pervasively permeate what is said by an utterance. Semantics is not autonomous from pragmatics – in some extreme versions, the contextual dependency is so intense that it is useless to talk of stable full-fledged conventional meanings.

Metaphor was seen in the early versions of the theory of relevance as a form of loose talk; more recent contextualist approaches propose that, when we interpret a metaphorical utterance, we modify the semantically encoded concept, constructing a new ad hoc concept – guided by the semantically encoded concept and by the context. This ad hoc concept would be part of what is directly expressed by the utterance, not only of what is merely communicated (Camp, 2006; Carston, 2002). The contextualist proposal manages to save a minimal version of compositionality – the meaning of the metaphor would still be computed compositionally, although using the ad hoc concept in those computations –, at the cost of allowing pragmatics to infiltrate into semantics.

Inferentialism, as developed by Robert Brandom (1994), is another alternative semantic theory which also defends the subordination of semantics to pragmatics. It would be interesting to study where inferentialism stands in the debate between traditional semantics and contextualism – in particular regarding the content of metaphoric expressions. Even if inferentialism has often been confronted with compositional standard semantics, its connections and conflicts with contextualism have been far less explored.

According to inferentialism, semantic notions are introduced in order to explain the pragmatic significance of linguistic performances. As the minimum linguistic move with pragmatic force is the utterance of a complete sentence – only by uttering a sentence can we perform a speech act –, content will be assigned primarily to full sentences. In particular, inferentialism holds that the meaning of a claim is determined by its inferential relations with other claims. The property of the inferential moves relating claims would be determined by the norms governing the discursive

practice. Isolated subsentential expressions cannot maintain inferential relations, so it would seem that we cannot attach any content to them.

Inferentialism, then, endorses both the semantic primacy of the propositional – as opposed to the traditional sub-sentential semantics – and semantic holism. It has been argued that this is not compatible with compositionality, and that therefore inferentialism is implausible. However, the productivity of language and the most relevant aspects of compositionality can be accommodated in an inferentialist framework. In order to see how, we need to look at the inferentialist reconstruction of subsentential semantics. The content of subsentential expressions would be determined in a derivative way, as the contribution these subsentential expressions make to the content of full sentences. Two subsentential expressions have the same indirect inferential role if whenever we substitute one for the other in a sentence, the inferential role of that sentence is preserved. We could characterise the indirect inferential role of a subsentential expression by looking at its substitutional relations with other expressions: how the inferential role of a sentence is affected when we substitute that expression for another one. Brandom describes a 'simple material substitutional commitment' between M and N as the commitment to accept a sentence in which N substitutes M whenever we accept the original sentence containing M (the inference from the sentence containing M to the sentence containing N is good). Then, the (simple material) substitutional commitments relating one subsentential expression to others would give us its inferential indirect content. We should stress again that these contents are derived indirectly from the full propositional contents/inferential roles.

The substitutional relations associated with singular terms are generally symmetric – in accordance with the intuitive idea that co-referential singular terms are symmetrically intersubstitutable. On the other hand, predicate's substitutional relations would be in general asymmetric: the substitutional inference from 'I am walking' to 'I am moving' is correct, but its reverse – from 'I am moving' to 'I am walking' – is not.

This subsentential semantic model explains some aspects of the productivity and creativity of language. Imagine that we know which are the asymmetric substitutional relations for a certain set of predicates; then, if a new singular term is introduced in the language, we will be able to understand many novel sentences involving this term, because we can already follow the substitutional implications of predicates. Therefore, the members of the linguistic community will be able to produce and understand sentences never used before – they will be able to confer pragmatic significance to moves that had never been made before in the practice.

In this way, inferentialism manages to incorporate a certain degree of compositionality. But, how does inferentialism cope with metaphor and with the contextualist challenges to traditional semantics? In this paper, I want to suggest that inferentialism offers a suitable framework for dealing with metaphor, as long as we allow for the possibility that the substitutional relations of expressions are contextual – in the sense that in some contexts of use an expression would have substitutional relations that do not coincide exactly with the relations hold in other contexts. Strict compositionality tries to solve the problem of the projectivity of language use by establishing a single uniform way of projecting – dictated by the fixed meanings of components and rules of combination –; but we might need a more fine-grained analysis, in which projections would depend on the context of use.

Such context sensitivity would make it possible for inferentialism to be receptive to some of the arguments posed by contextualists. At first sight, it may be thought that contextualism is incompatible with inferentialism because, according to the latter, the meaning of a claim would be given by a norm governing its use in an inferential practice. But the conflict would only arise if we demand that such semantic norm can be expressed by a general rule, which would specify the appropriate uses of the claim in a uniform way, without taking into account the different contexts of use. However, semantic norms can perfectly be context sensitive: what counts as a correct application of the norm would depend on the context of application. Explicit formulations of rules may just give us general guidelines that would help us to follow the norm properly in each particular case – or, perhaps, they would relate the present norm to some other norm that we know how to apply. Grasping the meaning of a claim would consist in mastering its inferential norms of use – norms which maybe cannot be fully captured in an explicit formulation. At the level of sub-sentential semantics, this context sensitivity would be reflected in the fact that, even if we accept that the meaning of a sentence is completely determined by the semantic contribution of its components[1], these components would not contribute to the meaning of aggregates in the same way in every context.

The context sensitivity of inferential relations may also offer a promising account of some aspects of metaphorical talk. Metaphor has been thought of as a mechanism by means of which the inferential web associated with a concept is partially projected from a source domain to a target domain (Lakoff and Johnson, 1980). We could say, then, than in metaphoric uses of a predicate, not

[1] Some contextual effects, such as enrichment, may be accounted for without linking them to the semantic contribution of any particular sub-sentential expression.

all the inferential relations of the predicate will be mapped from the original non-metaphoric use. The predicate frame 'a is cold' would have different substitutional relations when applied to objects or when applied metaphorically to people. How these projections work would be determined by the norms that regulate the way the practice functions. These norms would reflect an underlying conceptual metaphor (or a metaphorical schema), this is to say, a relation of mapping between the inferential structures of two topics or domains. Note that in this mapping what is projected is the structure of the inferential relations, not the inferential relations themselves. Take, for instance, the conceptual metaphor EMOTIONS ARE TEMPERATURE. When talking about the temperature of physical objects (source domain), we would say things like: 'That object is very cold, so it won't transmit thermal energy.' However, if the topic is human emotions (target domain), we will say instead: 'That person is very cold; she doesn't transmit affection.' The inferences are different, but their structure is related. When we learn the conceptual metaphor, we know that "affection" occupies in the inferential web of the target domain an analogous relative position to that which the concepts "thermal energy" or "heat" occupy in the source domain. That knowledge allows us to map inferential relations from one domain to the other.

This model of structural mapping accounts for metaphors in which no property is shared by the target and the source – something that poses a problem for theories of metaphor based on similarities (Carston, 2002). Cold people and cold objects do not share any relevant property; nevertheless, it is possible to establish a metaphorical connection between those two domains because there is a structural compatibility or correlation between the inferential webs by means of which we organize our discourse about them.

The proposal I have just suggested is an inferentialist version of the conceptual theory of metaphor (advanced by Lakoff and Jonshon, 1980). However, that latter theory pays attention mainly to the functions of metaphor in our cognitive processes, whereas inferentialism, as a theory of meaning, will be specially interested in the place of metaphor in our discursive practices – in metaphor as a linguistic phenomenon. If a conversational metaphor expresses certain semantic content, it will be because the mapping relations that underwrite the metaphor are mirrored by the semantic inferential norms governing the discursive practice. So, imagine that there is a conceptual metaphor in virtue of which we think of people who do not express their emotions as "cold people". This metaphorical way of thinking would be expressed by the semantic content of a claim such as "Robert is a cold person" only if there is a semantic norm according to which the inference from "Roberst is a cold person" to "Roberts doesn' t express his emotions" is a good inference.

Nevertheless, in some non conventional metaphors the projection pattern won' t be codified by the inferential norms of the language; then, the metaphor would lack a definite semantic meaning. In some cases, this kind of metaphors will be used in order to communicate indirectly some speaker's meaning – as Gricean pragmatists contend. It may also be the case that creative metaphors are uttered without a concrete communicative intention – there would be no clear propositional content that the metaphor tries to convey, either directly or indirectly. The metaphor would just invite the audience to explore the possible projections between the source and the target domain – in line with Davidson's proposal. I suspect that this is what happens with most poetic metaphors.

It could be argued that this proposal fails to offer a unified theory of the linguistic role of metaphors (Camps, 2006) – since we would attribute semantic content to some metaphorical utterances but not to others. However, nothing forces us to look for a homogeneous account of metaphor, and actually it may be advantageous to have a flexible theory, so we can cover the diverse uses of metaphoric discourse in human language and thought.

Another possible objection is that, once the inferential relations that articulate the target domain become registered in the discursive norms of the practice, the claims belonging to that domain lose their metaphorical character – they become dead or lexicalized metaphors. Indeed, it would seem that, if the inferential relations of claims in the target domain are regulated by standard semantic norms, the content associated to those claims would be just ordinary literal meaning; then, rather than talking about metaphorical concepts and meanings, maybe we should talk of polysemes[2]: a single term which expresses different literal meanings. The fact that those inferential relations emerged originally by means of a metaphorical mapping from a source domain, may be of interest for studying the history of the language, and perhaps its learning and cognition, but not for a theory of meaning.

Actually, this view would be natural if we think of normative practices as setting in advance a definite and fixed pattern of correctness, which then is (or ought to be) reproduced by those who follow the norm. Nonetheless, an alternative conception of normativity (influenced by certain readings of the latter Wittgenstein) holds that norm following is in a sense open-ended, and that in each new application of the norm, the participants of the practice keep constructing on the go the pattern of correctness (through their performances and the assessment of the performances of others). The correctness of new applications may not be determined beforehand by past uses – or by

[2] Or even of homonyms: different terms with different meanings, but sharing their lexical form (spelling and pronunciation).

explicit formulations of the norm. In the case of conventional metaphors, what would guide us when finding out how to follow the norm in new applications, would be a metaphorical projection from a source domain – whose semantic norms we are familiar with. Learning the metaphorical concepts pertaining to the target domain, would consist in learning to exploit those metaphorical projections in order to figure out how to follow the relevant semantic norms properly. A claim would be regarded as metaphorical, then, when our grasp of the correctness of its (new) inferential uses is derived from a metaphorical mapping. If we don' t need to resort to such metaphorical mapping, the metaphor would be genuinely dead – the metaphorical mapping would have ceased to be active.

There are more difficulties that should be addressed in time. Nevertheless, it seems worthwhile to develop in more detail an inferentialist account of metaphor along the lines of what has been presented above. In doing so, it is not only hoped that we will get a better understanding of metaphor, but also of inferential theories of language – and, in particular, of their engagement with contextualist arguments.

REFERENCES

Brandom, R. 1994, *Making It Explicit,* Cambridge, Harvard University Press.
Camp, E. 2006, 'Metaphor, contextualism and What is Said.', *Mind & Language,* 21:3, 280–309
Carston, R. 2002, 'Metaphor, *ad hoc* concepts and meaning – more questions than answers', *UCL Working Papers in Linguistics* 14, 83-105
Lakoff, G. y M. Johnson 1980, *Metaphors we live by*, Chicago, Chicago University Press.

DISPOSITIONS, AFFORDANCES AND MENTAL INTENTIONALITY[1]

MANUEL HERAS ESCRIBANO & MANUEL DE PINEDO
Departamento de Filosofía I, Universidad de Granada

ABSTRACT: In this paper we defend a bottom-up approach for explaining the ontological grounds of perception and cognition. We start by exploring some consequences of defending a realist conception of dispositions and we accept the thesis that dispositions display 'physical intentionality' (PI). Then we show how PI fits within ecological psychology's understanding of perceptual processes. Further, we defend a normative approach to the mental by which mental intentionality is nothing but rule-following: from a normative perspective there is a small set of dispositions accepted by the community, and this set is what shapes the life and practices of the individual. Finally, we will suggest that an understanding of normativity is enough for understanding consciousness.
KEYWORDS: Dispositions, affordances, physical intentionality, consciousness, ecological psychology, normativity

Within the debates on the naturalization of meaning, rationality or mental content it has been almost universally assumed that intentionality is an exclusive mark of the mental and the linguistic. The task of naturalizing anything normative would involve, according to this assumption, a naturalization of intentionality. To conceive of intentionality as a much more pervasive feature in nature would, at the very least, simplify the job for the naturalist.

In this paper we will remain neutral concerning both the feasibility of naturalizing the realm of norms and whether language is necessary for normativity. Instead, we will explore some consequences of defending (with Georg Molnar and C. B. Martin) a realist conception of dispositions accompanied by the thesis that they display 'physical intentionality'. In the process we will argue for several thesis: (1) the concept of 'affordance' is a good template for dispositional realism; (2) *mental* intentionality is unique because it is rule-bound, not because of its 'aboutness' or 'directness': normativity, rather than intentionality, is the mark of the mental; (3) mental, normative intentionality implies consciousness: only creatures aware of the possibility of error can

[1] Work for this paper was partially funded by the research project "Disposiciones, holismo y agencia" (Ministerio de Ciencia e Innovación, España, FFI2010-19455).

follow rules.

Molnar claimed that the four features that Brentano associated with the intentionality of the mental could also be found, with very small modifications, in dispositional properties. The features are: (i) an item with physical intentionality is directed towards something beyond itself; (ii) it could be directed towards something nonexistent – that is, something that happens not to be available; (iii) it is disposed towards exemplars and prototypes rather than specific items – it is not geared to something singular but to anything that fits a bill; and (iv) it is sensitive to the way the intended item is presented – the mode of presentation is relevant. Molnar 2003 and the papers collected in Martin 2008 have already become classic statements of this approach. John Heil (2003) and U. T. Place (1996) also defend variations of this idea.

For instance, a frog's disposition to eat flies is (i) directed towards flies, (ii) it continues to exist even in an environment without flies, say, during winter, (iii) it doesn't favour one specific fly but it would be satisfied by any fly that comes its way, and (iv) a fly-looking object would trigger it, while a pellet containing the nutrients of a fly would not.

While the defence of physical intentionally by Martin, Place, Molnar and Heil involves the idea that dispositions are directed towards their manifestations, our characterization and examples take dispositions to be directed towards their reciprocal dispositional partners. We are aware that this shift of focus is insufficient for arguing in favour of a relational conception of dispositions: dispositions may never encounter their partners or these may not even exist. However, thinking of dispositions as directed towards their triggers is a move in the direction of a more holistic conception of nature, one that is especially suitable for biological thinking.

Ecological psychology provides us with an attractive framework to explain how agents deal with their environment. Among the concepts developed within the field, the concept of 'affordance' is key to understand how an agent is capable of adapting itself to a changing ecosystem: affordances are opportunities for action, and for those opportunities to be taken it is necessary both an environment and an agent (Gibson 1966, 1979). When an agent actively exploring its environment detects specific information this means that there are specific features of the environment that demand an action; for example, the fly-looking object has features that invites the frog to catch it. This information complements the action of the agent and also modulates it: the selective history of the animal leads the frog to behave in that way for adapting to its environment. Also, these features are information at the ecological scale because of its guiding the action for the frog. This action

generates a new action and it allows the agent-environment loop to be constant (Richardson *et al.* 2008).

The ecological scale is a non-reductive or eliminative level of explanation to which ecological psychology is committed. Instead of explaining the mark of the cognitive within physics, ecological psychology demands a new realm for explaining the cognitive and intentional: this perspective is the ecological scale, in which the right level of explanation of the cognitive comes from the dynamic interactions between agent and environment. We said that the animal detects 'specific information': this is information for performing an action. This information is a feature of the environment which exerts selection pressure to the animal; it forces the animal to adapt itself to this new aspect of its habitat. This pressure transforms an inert informational stimulus into a meaningful one. When the frog detects visual information of the presence of a fly (meaningful information demanding an action), the frog tries to catch the fly. This information is of a special kind: it is not physical information, but information related to the level of agency, action and behaviour. The frog does not detect *just* atoms, particles, etc. What the frog detects is a compound of textures, colours, shapes, and shadows that carry the information that the object is edible. So, the analysis of the behaviour of the animal as an answer to the meaningful stimulus of the environment understood as a continuous, dynamical interaction at the level of the agent is the ecological scale. This explains the particular phenomenon of affordances, and provides a mark for applying the notion of physical intentionality mentioned above.

Note two consequences of applying the notion of physical intentionality to affordances: first of all, as we have said, affordances are the result of the adaptive behaviour of the animal to its environment; second, there is no possibility of error: affordances depend both on properties of the animal and information at ecological scale; this is, information *related to* an animal. This is why there can be no error: when it recognizes an informative feature of the world, its reaction is the result of an intentional action determined by its evolutionary history. If an evil ethologist decides to change flies for pellets in an animal's environment with no pellets before, its decision says nothing of the behaviour of the frog. The animal has done its best to regulate its action and adapt to its environment, and its behaviour is the right one at ecological scale.

According to most of specialized authors, that is the minimal basis for perception and it requires of two elements to be achieved. On the other hand, for most ecological psychologists affordances are both (1) ecological (because the minimal unit of analysis is the Organism-Environment System) and also (2) a real fact of the world (which means 'agent-independent').

Then, there seems to be a conceptual tension: How can an affordance be both 'ecological' (agent-dependent property of the environment) and 'real' (agent-independent aspect of a habitat)? That is why the ontological status of affordances is a relevant question in the realm of ecological psychology. At the end, most efforts for explaining the ontology of affordances have been considered unsatisfactory because of different reasons.

We will focus on the three main theories of affordances (Turvey, 1992; Reed, 1996; Chemero, 2009) and we will show how at the end of the debate there is a trade-off: some authors defend the ecological scale as the ontological mark of the affordance instead of their real status, and some others support the opposite option.

Our proposal is to show why this trade-off is not the only solution. For this, we will apply a realist conception of dispositions and will make use of Martin's (2008) idea of reciprocal dispositional partnership to ground such processes. We claim that the very notions related to dispositional realism (readiness, directedness, etc.) are available for explaining the elements and requirements for perceiving as they are understood in the field of ecological psychology. As we have said, if the basic unit of analysis is the Organism-Environment System, we can understand this basic unit as the combination of two reciprocal dispositional partners whose directedness relates them to each other. Both organism and environment, when coupled, *can* manifest the act of perceiving; at the same time, the manifestation of these properties is not what makes them real: an unmanifested affordance does not mean that there cannot be properties ready to go in both sides – properties that, even unmanifested, are real. This approach thus preserves both the realist requirement (because those properties exist in the environment and also in the organism independently of their manifestations) and the ecological scale at the same time (both properties are explained by ecological psychology, and this means that they conform the cognitive-behavioural level of the agent).

This realist, dispositional approach can provide a new way to clarify the very notion of affordance and dissolve the trade-off for ecological psychologists. We also offer a final consideration on how this dispositional approach is related to a new way of understanding other kind of processes: mental intentionality and consciousness.

Concerning the former, if we go one step further than mental perception, (some) animals are supposed to display mental intentionality. Traditionally, mental intentionality has been linked to language in some strong or weak relationship (usually explained in terms of sufficient or necessary

conditions). As we said, we will remain neutral in that respect. That means focusing not just on linguistic agents, but on any agent capable of knowing and with a sufficient flexible and social behaviour in its ecosystem.

Based on this, we defend a normative approach to the mental by which mental intentionality is nothing but rule-following (Wittgenstein, 1953), although this does not mean that there is no trace of dispositionality in this cognitive-behavioural level (Ryle, 1949). All animals show directness and aboutness, but some of them are also able to follow rules and to learn skills within their community. Normativity is different from intentionality because for the latter there can be different dispositions towards some objects, real or not, and all of them are equally valid; for the former, just some of those are the right ones: from a normative perspective there is a small set of them accepted by the community, and this set is what shapes the life and practices of the individual. That is, from a Wittgensteinian perspective, to be immersed in a form of life in which repetition, regularity, training, and recognition of the possibility of error bound all skills and practices. And that is already a step beyond the mere directedness or aboutness of non-social agents, because a practice is not an isolated disposition directed to an object, but a trained disposition with correction and evaluative criteria.

Related to the possibility of error, we consider that to recognize mistakes and to module one's behaviour in order to avoid them is enough for being conscious of something (while knowing that one has a perspective on things that sometimes fails to do justice to them). While we can just talk of 'awareness' for physical intentionality (because an animal, when perceives something, is directed to the detection of some kind of information and this makes it aware of this directedness), we should talk of consciousness when we enter in the realm of the normative. Anyway, both awareness and consciousness should not be considered as some extra phenomenon that anyone has to find or localize in the agent: the traditional split between intentionality and consciousness is not that clear when we accept that consciousness is nothing but knowing that one is able to be in a state of error when putting into practice any skill. Thus, to say that somebody is 'conscious' is nothing more that acknowledging the rational behaviour of that individual within the community: somebody who is capable of being mistaken and also of being right and to correct her errors. This is so because 'to be mistaken' implies to be mistaken *about something*, and this means that one already has the concept of something and also that she is able to distinguish between herself and states of herself, on the one hand, and the world and states of the world, on the other. Consciousness (being aware of acting under correction criteria) and self-consciousness (being aware of one's own

thoughts) go hand in hand with mental, normative intentionality.

In conclusion, we want to prove that intentionality can be found in all processes in nature because of its intrinsic directedness. When we talk of animals in their ecosystem, it is unavoidable to talk of affordances as the key phenomenon for explaining perception; but when we talk about the mental, normativity enters the scene. To sum up all these processes in a picture, the most suitable framework is dispositionality: for both physical and mental intentionality, dispositions are the best conceptual apparatus to explain the continuity from world to mind within a naturalistic perspective.

REFERENCES

Chemero, A. 2009, *Radical Embodied Cognitive Science*, Cambridge, Mass., The MIT Press.
Gibson, J. J. 1966, *The Senses Considered as Perceptual Systems*, Boston, Houghton Mifflin.
Gibson, J. J. 1979/1986, *The Ecological Approach to Visual Perception*, Boston, Houghton Mifflin.
Heil, J. 2003, *From an Ontological Point of View*, Oxford, Clarendon.
Martin, C. B. 2008, *The Mind in Nature*, Oxford, Oxford University Press.
Molnar, G. 2003, *Powers: A Study in Metaphysics,* Oxford, Oxford University Press.
Place, U. T. 1996, *Dispositions: A Debate* (Armstrong, Martin and Place, edited by Tim Crane), London, Routledge.
Reed, E. S. 1996, *Encountering the World*, Oxford, Oxford University Press.
Richardson, M. J., Fajen, B. R., Shockley, K., Riley, M. A., Turvey, M. T. 2008, "Ecological Psychology: Six Principles for an Embodied–Embedded Approach to Behavior", in P. Calvo & T. Gomila (eds.), *Handbook of Cognitive Science: An Embodied Approach*, Amsterdam, Elsevier, 161-197.
Ryle, G. 1949, *The Concept of Mind*, London, Hutchinson.
Turvey, M. 1992, "Affordances and prospective control: an outline of the ontology", *Ecological Psychology*, 4(3), 173-187.
Wittgenstein, L. 1953, *Philosophical Investigations*, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.

CONCEPT-ASPECTS AND THE SPECIFICATION OF COGNITIVE PHENOMENOLOGY[1]

MARTA JORBA-GRAU
LOGOS- Universitat de Barcelona
martajorba@ub.edu

ABSTRACT: When we consciously think a thought or entertain a certain proposition we undergo a certain experience. The phenomenal character involved in such an episode is an instance of cognitive phenomenology, at least in the sense that there is *some* phenomenal character in the episode of conscious thought. Controversies arise with respect to the *nature* of such cognitive phenomenology, whether it is specifically cognitive or it is reducible to more familiar kinds of phenomenologies, such as the sensory or emotional one. In this paper I examine an objection against the specificity thesis, based on a fundamental asymmetry between the cognitive and the perceptual/emotional domain and I sketch a a proposal for the specification of cognitive phenomenology that responds to it and at the same is a defense of a specific cognitive phenomenology.
KEYWORDS: cognitive phenomenology, concepts, sensory phenomenology, reductionism.

When we consciously think a thought or entertain a certain proposition we undergo a certain experience. We can say that the phenomenal character involved in such an episode is an instance of cognitive phenomenology, at least in the sense that there is *some* phenomenal character in the episode of conscious thought. Some agreement is reached concerning the *existence* of cognitive phenomenology, when this thesis is not filled in with more substantial claims (see Smithies forthcoming). The mere existence of an experience of consciously thinking is not problematic *per se*, but controversies arise with respect to its *nature.* A question we need to answer is whether cognitive phenomenology is *specifically cognitive* or it is *reducible* to more familiar kinds of phenomenologies, such as the sensory or emotional one2.

Some authors deny any progress in the specification of cognitive phenomenology on the basis of a fundamental asymmetry with the perceptual (and emotional) case: in the case of conscious thought we do not find any commonalities and differences with respect to phenomenal character that would secure speaking of experiential kinds (Martin manuscript). Georgalis (2005)

[1] Financial support for my work was provided by the DGI, Spanish Government, research project FFI2010-16049 and by the FPU Program of the Spanish Ministry of Education and Science.
[2] For an overview of the problem, see Bayne, T. & Montague, M. (2011).

further argues that this is an obstacle to render cognitive phenomenology essential to conscious thought. This objection can be turned into a challenge for the proponent of cognitive phenomenology: if one could provide an account of the relevant salient features of the phenomenology of conscious thought, we would have reason to think that the objection is misguided and we would have made some progress towards the specification of cognitive phenomenology. In this paper, thus, I sketch an account that can overcome this objection in a way that also gives evidence for the specificity of cognitive phenomenology. In order to be able to assess this solution, let me briefly present the main lines of the debate between defenders of a specific cognitive phenomenology and reductionists.

Some arguments regarding cases for specific cognitive phenomenology[3] normally have the form of phenomenal contrast argument: they present two scenarios where there is a phenomenal change from one to another and nevertheless the non-cognitive components (mainly sensory and perceptual aspects) remain the same. Since, it is argued, the only difference between both scenarios is cognitive, the phenomenal contrast is to be accounted for by appealing to a specific cognitive phenomenology. The paradigmatic case is that of *understanding* vs. *not understanding* some written or heard sentence, for example (Strawson 1994/2010). Or cases of the experience of comparing the price of some item in a familiar currency versus comparing it in a foreign one.[4] There are also single cases: the example of *imageless thought*, when thoughts are conscious but nevertheless lack images[5] or the cases of *languageless thought,* where there is some consciousness of the thought without any sentence being experienced, like sudden thoughts that occur without time for language (Siewert 1998).

The authors on the reductionist side of the debate[6] try to accommodate all these cases by postulating differences in some more familiar kinds of phenomenal character, like the sensory or emotional one. A very extreme view on this respect is Prinz (2007, 2011), who argues for the claim that all consciousness is *perceptual consciousness*[7]. He tries to accommodate the phenomenal

[3] Provided by Husserl (1900/1901); Kriegel (2011); Peacocke (1998); Pitt (2004); Siewert (1998); Strawson (1994/2010).

[4] Thanks to Manuel García-Carpintero for this example.

[5] The discussion on imageless thought goes back to introspective psychology and the debate whether imageless thoughts are possible. The experiments where based in people reporting whether they had images or not when asked certain questions: what substances are more costly than gold, etc?. These experiments and introspectionist psychology was after that highly dismissed and regarded as a failure of method: subjects can simply be wrong about their own mental states (see Prinz 2011, for an overview of the debate).

[6] Carruthers, P. and Veillet, B, 2011; Prinz, 2007, 2011; Tye and Wright, 2011, among others.

[7] Prinz argues for the particular claim that all *phenomenal* consciousness is perceptual phenomenal consciousness, and he believes other uses and forms of consciousness are parasitic on phenomenal consciousness, and thus this more general claim can be defended (see Prinz, 2007, p. 336). This view constrast, for example, with Peacocke's (), according to which conscious thought is a special case of another kind of conciousness, namely, action

contrast of understanding and similar cases with differences in sensory elements, such as different associated mental images or inner speech or differences in the focus of attention. He then allows for *conscious* thoughts as long as there are no qualities over and above the sensory ones: there is an acceptance that conscious thought "feels like" something (there is a phenomenology), but not that it feels differently than sensory activity (all phenomenology is reducible to sensory one).

The account I am putting forward here is a proposal for the specification of cognitive phenomenology and at the same a defense of a specific cognitive phenomenology. It is based on two aspects of concepts, the world-tied element and the conceptual network (further specified in relation to background knowledge): while the first provides us with commonalities in cognitive phenomenology, the second, together with sensory/emotional elements, accounts for its differences. Consider the following example. You and me are standing in front of a field and we both see a flower and think the proposition that the flower is beautiful. What are the *experiential commonalities* and *differences* between me and you in the cognitive state? The sense in which we have the *same* kind of thought can be explained by the fact that our thoughts are about the same things, namely, the flower stands in front of us. A familiar picture of why this is so is the fodorian picture of concepts as concrete mental particulars in the head that are constitutively linked to the world and thus externally individuated. If these world-tied aspects of concepts have any contribution to phenomenology, I think they provide us with *commonalities* in cognitive phenomenal character, or at least it is not *in virtue of* the world-tied aspect of concepts that we have different cognitive experiential mental states. The differences in phenomenal character are provided by differences in *inner speech* (yours being a certain kind of tone and speed and mine another), *images* associated with this thought and *possible* feelings associated with it. If we suppose, for the sake of the argument, that all these elements are the same in you and me, as are the concepts FLOWER and BEAUTIFUL externally individuated, can we still say that you and me have different experiential or phenomenal character while entertaining the thought that the flower is beautiful?

Against cognitive phenomenology reductionists[8], I argue that the answer is yes. The sensory, perceptual and emotional elements in the episode of conscious thought do not suffice to explain the differences between you and me in the phenomenal character of the conscious thought, precisely because of the *network* in which our thoughts are embedded in our mental economies. My proposal

consciousness. Action awareness is the other case apart from thought than can provide objections to Prinz's view (Prinz, 2007, p. 341).

[8] They would predict that if there are no sensory, perceptual, emotional phenomenal aspects, no phenomenal difference can be in place, as these are all kinds of phenomenologies we should acknowledge.

is that the connections and the "situation" of the concept in our cognitive mental lives form a kind of network, that is needed to account for further cognitive experiential differences, over and above the elements mentioned. This network is constituted by the *background knowledge* one possesses about a certain concept and that we both may differently carry. The idea is that the more knowledge one has over a certain subject, the bigger the network is, and more differences in cognitive phenomenology we can find or the richer it is. The connections of this network are clearly not differences in sensory phenomenology, so if my proposal is sound, reductionists positions of the sort presented are in trouble.

One might object that we are talking about *occurrent* conscious thought, and the network is a set of dispositional concepts that cannot be experientially present when we entertain the thought, so they are elements that cannot account for the cognitive phenomenal difference between you and me. This proposal has to be developed in more detail, but the idea and response to this objection would be that the network is somehow felt with the occurrent concept you are thinking about, just in the sense in which one can say that there is phenomenal consciousness in the peripheral areas of the visual field that are not the focus of our attention.

How does the sketched view help to answer the objections from Martin and Georgalis? I think it undermines the objection that in conscious thought we do not find salient phenomenological similarities and differences, as it shows how they can be traced to different aspects of the concepts, the world-tied aspect and the network one: while the former may be responsible for the commonalities we find in the cognitive experience of a conscious thought, the latter, which is intimately related to and influenced by background knowledge, may account for the experiential differences over and above sensory ones. These elements may generalize as to provide us with experiential kinds. This sketched view could thus provide a first step towards a specification of the phenomenology of conscious thought and at the same time is a defense for the claim that there is a specific cognitive phenomenology.

REFERENCES

Bayne, T. & Montague, M. 2011 (eds.), *Cognitive Phenomenology,* New York, OUP.

Bayne, T. and Montague, M. (eds.). 2011, "Cognitive Phenomenology: and Introduction", in Bayne, T. and Montague, M. 2011, *Cognitive Phenomenology,* New York, OUP.

Georgalis, N. 2005, *The primacy of the Subjective: Foundations for a Unified Theory of Mind and Language.* Chapter 3, MIT Press.

Husserl, E. 1900-1901/1970, *Logical Investigations*. London and New York, Routledge.

Kriegel, U. forthcoming, *The Varieties of Consciousness. Studies in non-sensory phenomenology.*

Martin, M. manuscript, *Uncovering Appearances.*

Peacocke, Ch. 1998, "*Conscious attitudes, attention, and self-knowledge", in Wright, C., Smith, B.*

and MacDonald (eds.). Knowing our own minds, OUP.

Pitt, D. 2004, "The phenomenology of cognition, or, what is it like to think that p? *Philosophy and Phenomenological Research*, LXIX, No. 1.

Prinz, J. 2007, "All consciousness is perceptual", in B. McLaughin and J. Cohen (eds), *Contemporary Debates in Philosophy of Mind,* Oxford, Blackwell, p. 335-57.

Prinz, J. 2011, "The Sensory Basis of Cognitive Phenomenology", in Bayne, T. and Montague, M. (eds). 2011, *Cognitive Phenomenology,* New York, OUP.

Siewert, Ch. 1998, *The Significance of Consciousness*. Princeton University Press.

Siewert, Ch. 2011, "Phenomenal Thought", in Bayne, T. and Montague, M. (eds)(2011). *Cognitive Phenomenology,* New York, OUP.

Smithies, D. forthcoming, *"*The nature of cognitive phenomenology", in *Philosophy Compass.*

Strawson, G. 1994/2010, *Mental Reality,* 2nd ed. Cambridge, MA: MIT Press.

Strawson, G. 2011, "Cognitive phenomenology: Real life", in Bayne, T., and Montague, M (eds). *Cognitive Phenomenology,* New York, OUP.

Tye, M., and Wright, B. 2011, "Is there a phenomenology of thought?", Bayne, T., and Montague, M (eds). *Cognitive Phenomenology,* New York, OUP.

DECISIONES EPISTÉMICAS Y LOS LÍMITES DEL NATURALISMO[1]

MANUEL LIZ
Universidad de La Laguna
manuliz@ull.es

ABSTRACT: The paper analyses an argument according to which the existence of epistemic decisions would pose a crucial problem for naturalism. The argument is the following one. We could adopt a naturalistic perspective in many fields. We could adopt such a perspective even in fields such as those of morality and action. However, we could not adopt a naturalistic perspective with respect to the fact that we are able to adopt epistemic decisions. To adopt a naturalistic perspective with respect to our epistemic decisions would preempt the very fact of adopting such a perspective. Epistemic decisions have to necessarilly incorporate a certain feeling of spontaneity that would be cancelled if a naturalist perspective were to be adopted. We distinguish a direct version and a modalized version of the argument. And we construe a thought experiment to illustrate the situation generated by them. In that thought experiment, the condition of being like an "epistemic puppet" is contrasted with the condition of being like an "agentive, evaluative and normative puppet".
KEY WORDS: Epistemology, epistemic decisions, naturalism, arguments against naturalism, spontaneity, blind mechanisms, transcendental arguments.

En mi trabajo, defenderé la tesis de que nunca podríamos llegar a adoptar una perspectiva naturalista completa sobre nosotros mismos. El argumento es tan simple como contundente. Una tal perspectiva naturalista completa sería literalmente imposible de "adoptar". Adoptarla anularía automáticamente lo pretendido en dicha "adopción". Dicho de otro modo, la cancelaría.

Se han esgrimido muchos argumentos contra los proyectos naturalizadores, especialmente en los ámbitos relativos a la epistemología y a la filosofía de la mente. Algunos de ellos son ya clásicos. Los proyectos naturalizadores han sido repetidas veces acusados de cometer algún tipo de letal falacia naturalista respecto de los componentes normativos y evaluativos de los fenómenos que se intentaban naturalizar. También han sido acusados de presuponer ilegítimamente que el conocimiento era posible, y que podían tranquilamente dejarse al margen los escenarios escépticos cartesianos. También se ha considerado que nuestras capacidades reflexivas generan

1 Este trabajo ha sido llevado a cabo en el seno de los proyectos de investigación FFI 2008-01205 y FFI 2011-24549.

inevitablemente explosiones de nuevos fenómenos, intencionales o cualitativos, imposibles de tipificar en términos naturalistas.

¿Cómo hacer frente a estos argumentos desde una perspectiva partidaria de la naturalización? Sin duda, la mejor estrategia consiste en subrayar que ninguno de ellos es definitivo. Las explosiones de nuevos fenómenos intencionales o cualitativos generados en la reflexión no tienen porqué ser radicalmente diferentes de los fenómenos ya tipificados en términos naturalistas (véanse las discusiones al respecto de Sosa 1981 y 2009). Por otro lado, presuponer que el conocimiento es posible, y que pueden arrinconarse las posibilidades escépticas radicales, no es una objeción específica contra el naturalismo. Cualquier presunto conocimiento de la realidad podría verse igualmente deslegitimado por una objeción similar (lo cual, por cierto, serviría de contrapunto a las reiteradas críticas de Stroud, 1981, a Quine, 1969). Y la comisión de una falacia naturalista podría dejar de ser tan letal cuando la naturalización se formula a través de relaciones muy generales de sobreveniencia (véase al respecto Kim, 1988).

El argumento que quiero discutir tiene mucha más fuerza que todos los anteriores. ¿Por qué la existencia de decisiones epistémicas entraña serios problemas para cualquier proyecto naturalizador? La respuesta es ésta: adoptar una decisión epistémica siempre requiere ciertos rasgos cualitativos que no pueden estar presentes si asumimos que esa decisión no es más que el resultado de algo que podríamos describir como una serie de “mecanismos ciegos”.

Ahora bien, ¿qué puede significar aquí ser un “mecanismo ciego”? Podemos decir que ser un “mecanismo ciego” es ser una entidad perteneciente, aunque sea en un sentido muy genérico, a alguno de los tipos de entidades que usualmente encontramos en el mundo natural tal como es descrito por nuestra ciencia.

Los rasgos cualitativos relevantes tiene que ver sobre todo con el sentimiento de “espontaneidad” que parece ser un ingrediente necesario de toda decisión epistémica. Podemos asumir pérdidas de ese sentimiento de espontaneidad en muchos campos, podemos incluso asumir en muchos casos pérdidas totales, pero no podemos asumir una pérdida total de la espontaneidad en el campo de las decisiones epistémicas.

Podemos llegar a admitir que todas nuestras decisiones morales, políticas, estéticas, etc., así como todas nuestras decisiones gastronómicas, laborales, sexuales, etc., son el resultado de “mecanismos ciegos”. Sin embargo, no podemos admitir que todas nuestras decisiones epistémicas sobre qué creer y qué no creer entren en el mismo saco. Admitirlo es incompatible con tomar decisiones epistémicas. Si realmente existen decisiones epistémicas, no podemos adoptar respecto a

ellas una posición naturalista. Este es el argumento principal que quiero discutir. Podemos llamarlo el argumento directo.

Pero hay una ampliación modal importante del argumento anterior. En la medida en que "concebir" también involucre decisiones epistémicas, cabe argumentar que ni siquiera podremos concebir que realmente puedan existir decisiones epistémicas sin un genuino sentimiento de espontaneidad no cancelado. Ni siquiera podremos concebir que pueda adoptarse una perspectiva naturalista respecto a nuestras decisiones epistémicas.

Tanto el argumento directo como el argumento modal son importantes. Pero son distintos. El segundo es obviamente mucho más fuerte y comprometido que el primero. En cualquier caso, creo que son defendibles ambas versiones.

El naturalismo debería dar algún tipo de explicación robusta del sentimiento de espontaneidad que parece debe estar siempre presente en nuestras decisiones epistémicas. Ese rasgo debería quedar integrado en una cierta imagen naturalista del mundo. O al menos, el naturalismo debería indicar cómo sería posible esa integración. Aunque sea en términos tan generales como los de una relación de sobreveniencia respecto a los hechos naturales, Sin embargo, el naturalismo no puede hacer esto. Ofrecer cualquier explicación de ese tipo implicaría una pérdida irreparable de espontaneidad. En último término, la espontaneidad cualitativa de las decisiones epistémicas no puede ser considerada algo genuino, es decir no simplemente una ilusión o una fantasía, y a la vez algo superveniente a hechos naturales tipificables por una ciencia similar a la que actualmente tenemos.

Tal vez el problema desaparecería con una ciencia enriquecida de la manera sugerida por Sellars (1962). Pero en este tipo de enfoque, la espontaneidad tendría un carácter cualitativo no meramente epifenoménico o derivado. Su carácter cualitativo sería irreduciblemente básico y primitivo.

Las decisiones epistémicas se someten a todo tipo de constricciones externas. Algunas de esas constricciones son de tipo normativo e ideal. Otras son abiertamente naturales. En cualquier caso, tomar decisiones epistémicas intentando respetar ciertas normas de racionalidad, o ajustándose a ciertos parámetros adaptativos, o a ciertas condiciones de éxito pragmático, o acaso intentando seguir la pista de la verdad en un sentido confiabilista, deben ser cosas asumidas como algo que nosotros hacemos de manera espontánea. El detalle crucial es que no podemos asumir que sean cosas que simplemente "nos pasen a nosotros", o cosas que meramente "pasan" en un sentido absolutamente impersonal. (Sobre la necesidad de un ascenso epistémico en la justificación de

nuestras creencias, véase nuevamente Sosa, 1981 y 2009. Lo que quiero argumentar aquí es que ese ascenso epistémico ha de ser inevitablemente "personal".)

Tampoco podríamos adoptar aquí ninguna distinción entre un "nivel empírico" y un "nivel trascendental". No ser espontáneos en algún nivel trascendental también cancelaría el sentimiento cualitativo de espontaneidad que debe estar presente en toda decisión epistémica. La espontaneidad requerida debe darse en todos los niveles. Sugerir una "construcción trascendental" de nuestras decisiones epistémicas vale tanto, o mejor dicho tan poco, como sugerir que sólo son el resultado de "mecanismos ciegos". (Esta línea argumental es explorada en detalle por Nagel, 1997. Boghossian, 2006, también hace abundante uso de ella en sus críticas al relativismo.)

La idea de que las decisiones epistémicas, pero no necesariamente las decisiones morales, políticas, estéticas, etc., han de ir siempre acompañadas de un sentimiento cualitativo de espontaneidad está profundamente enraizada en nuestra autoconcepción como seres humanos. Para mostrar hasta dónde llega tal enraizamiento, utilizaré a continuación un sencillo experimento mental.

Supongamos que por alguna razón (tal vez sea la única manera de seguir vivos) estamos obligados a elegir entre o bien ser conectados a un dispositivo que controla todas nuestras decisiones epistémicas, dispositivo E, o bien ser conectados a un dispositivo que controla todas nuestras decisiones agentivas, evaluativas y normativas, dispositivo A. En el primer caso, nos convertiríamos en una especie de "marionetas epistémicas", en el segundo caso acabaríamos siendo una especie de "marionetas agentivas".

Suponemos que ambos dispositivos son suficientemente independientes. Aunque esto suponga una fuerte idealización, no importa demasiado. Se trata de un experimento mental diseñado para explorar el alcance de nuestros conceptos. En particular, tratamos de identificar una diferencia entre la noción de tomar decisiones epistémicas y la noción de tomar cualquier otro tipo de decisiones.

No sabemos cuál es realmente nuestra situación actual, y el argumento que estoy presentando es independiente de esto, pero sí sabemos dos cosas:

1. si nos conectamos al dispositivo E, a partir de ese momento todos nuestros sentimientos de espontaneidad epistémica no serán más que una mera ilusión, como mucho algo epifenoménico, y

2. si nos conectamos al dispositivo A, a partir de ese momento todos nuestros sentimientos de espontaneidad agentiva, evaluativa y normativa no serán más que una mera ilusión, o también algo epifenoménico.

Como dijimos antes, si escogemos ser conectados al dispositivo E, seremos meras "marionetas epistémicas". Si escogemos ser conectados al dispositivo A, seremos meras "marionetas agentivas, evaluativas y normativas".

Debemos decidir entre una cosa u otra. ¿Qué escoger? Creo que elegiríamos ser conectados al dispositivo A antes que al B. Preferiríamos antes convertirnos en meras marionetas agentivas, evaluativas y normativas, pero poder saber que lo somos, que convertirnos en meras marionetas epistémicas, y así no poder siquiera saber si somos o no somos meras marionetas agentivas, evaluativas y normativas.

En una palabra, al elegir el dispositivo B se pierde ciertamente mucho de lo que nos hace humanos. ¡Pero elegir el dispositivo A implica perderlo todo!

Valoramos mucho todas nuestras decisiones. Sin embargo, las decisiones epistémicas tienen un valor añadido. Consideramos más importante poder llegar a saber que somos meras marionetas agentivas, evaluativas y normativas, si es que acaso lo somos, que tal vez no serlo pero no poder llegar a saberlo nunca.

Aceptaríamos grandes pérdidas de espontaneidad en muchos ámbitos, pero no en el de nuestras decisiones epistémicas. Lo último que quisiéramos perder es esta clase de espontaneidad. Por ello, no se trata tan sólo de que la existencia de decisiones epistémicas suponga un serio problema para cualquier proyecto naturalizador. Esto es cierto. Pero la cuestión de fondo es que difícilmente podríamos vivir sin suponer que realmente adoptamos decisiones epistémicas.

REFERENCIAS

Boghossian, P. 2006, *Fear of knowledge. Against relativism and constructivism*, Oxford, Carendon Press.
Kim, J. 1988, "What is 'naturalized epistemology'?", en J. E. Tomberlin (ed.) Philosophical Perspectives 2: Epistemology, Atascadero, Ridgeview, 1988.
Nagel, T. 1997, *The Last Word*, Oxford, Oxford University Press.
Quine,W. 1969, "Epistemology Naturalized", en Ontological Relativity and Other Essays, New York, Columbia Univ. Press.
Sellars, W. 1962, "Philosophy and the Scientific Image of Man", en Science, Perception and Reality, London, Routledge and Kegan Paul, 1963.
Sosa, E. 1981, *Knowledge in Perspective*, Cambridge, Cambridge Univ. Press.
Sosa, E. 2009, *Reflective Knowledge*, Oxford, Clarendon Press.
Stroud, B. 1981, "The Significance of Naturalized Epistemology", en Midwest Studies in Philosophy, vol. VI, Minneapolis, Univ. of Minnesota Press.

"YO SOY GEORGE KAPLAN" TRES TESIS SOBRE LA RIGIDEZ SEMÁNTICA[1]

MANUEL LIZ
Universidad de La Laguna
manuliz@ull.es

ABSTRACT: The paper focuses on the semantical notion of rigidity. Hitchcock's film North by Northwest is used to argue for three theses: 1) Different sets of tokens of the same type of expression can refer in very different ways in approximately the same contexts of use, 2) reference is always determined by the referential intentions associated to tokens of some types of expressions and by certain mechanisms trying to stabilise some of the referential uses involving those tokens, and 3) neither rigidity concerning sets of tokens of certain types of expressions, nor rigidity concerning those types of expressions justify the introduction of modalities different from physical modalities and logical modalities.
KEY WORDS: Language, reference, semantical rigidity, tokens of expressions, types of expressions, metaphysical modalities.

Utilizaré la película de Alfred Hitchcock, de 1959, Con la muerte en los talones (el título original inglés es North by Northwest), para defender tres importantes tesis sobre la rigidez semántica. A veces, este fenómeno se ha entendido de una manera, por decirlo así, excesivamente rígida. Y la película de Hitchcock ilustra perfectamente algunos aspectos cruciales en los que deberían adoptarse posiciones mucho más flexibles.

Entenderé la noción de rigidez en su sentido habitual. Empleando la terminología de los mundos posibles, podemos decir que una expresión lingüística refiere rígidamente si refiere siempre a la misma entidad en todo mundo posible en el que dicha entidad exista. Sin duda, debemos a Saul Kripke (1972) la introducción de esta noción. Y está estrechamente ligada al enfrentamiento entre partidarios de la referencia directa, al menos para algunos tipos de expresiones, y partidarios de la necesaria mediación de la referencia por el sentido, principalmente en el sentido de "sentido" que lo equipara con alguna clase de descripción, o conjunto de descripciones. A su vez, las teorías de la referencia directa han estado muy vinculadas a diversas teorías de tipo causal-histórico acerca de cómo puede originarse y transmitirse tal referencia directa. No quiero discutir aquí ninguno de estos

1 Este trabajo ha sido llevado a cabo en el seno de los proyectos de investigación FFI 2008-01205 y FFI 2011-24549.

temas. Aunque algunas de mis observaciones tengan consecuencias relevantes respecto a la mayoría de estas polémicas, mi objetivo es analizar algunos aspectos de la propia noción de rigidez. También voy a intentar esquivar la mayoría de los problemas que con más frecuencia han suscitado las posiciones de Kripke. Tan sólo abordaré uno de ellos: el problema del tipo especial de necesidad metafísica que parece sugerir el fenómeno semántico de la rigidez.

Mi interés se centrará en el propio fenómeno de la rigidez: en cómo la rigidez semántica puede atribuirse de modo variable a diferentes clases de tokens de un mismo tipo de expresión lingüística, y en cómo puede atribuirse rigidez semántica a esos tipos de expresiones. Y como he dicho, el único de los temas kripkeanos que trataré es el de si en alguno de estos sentidos la rigidez justifica la introducción de un tipo especial de modalidades metafísicas. Mi respuesta será negativa.

Me apoyaré en la película de Hitchcock Con la muerte en los talones para defender varias tesis sobre la rigidez. El protagonista es un publicista de Nueva York. Su nombre es Roger Thornhill y el papel es interpretado por Cary Grant. Roger Thornhill es confundido con George Kaplan por parte de un grupo de espías extranjeros. Pero "George Kaplan" era el nombre de una entidad de ficción. Tal referencia ficcional es creada por los servicios de contraespionaje del gobierno. Estos habían construido la identidad de un agente ficticio con el propósito de desorientar a los anteriores espías. Supuestamente, George Kaplan ocupaba una habitación de un conocido hotel. Y desarrollaba todo tipo de actividades que eran minuciosamente seguidas por el grupo de espías. Para los servicios de contraespionaje del gobierno, el nombre "George Kaplan" refiere a una entidad de ficción. Para el grupo de espías extranjeros, sin embargo, "George Kaplan" refiere a través de un conjunto de descripciones. Los espías no tienen duda de que algunas de ellas pueden ser falsas. Pero esto no importa. A través de ellas, ellos quieren identificar al agente George Kaplan. Ciertas coincidencias desafortunadas hacen que los espías acaben creyendo que Roger Thornhill es George Kaplan. Le secuestran, le interrogan sin éxito, e intentan matarle. Él consigue librarse. Pero siguen persiguiéndole de forma implacable. Y entonces, algo sorprendente ocurre (algo, al menos, semánticamente sorprendente). En cierto momento de la película, Roger Thornhill acaba asumiendo la identidad de George Kaplan. Incluso manifiesta decididamente "Yo soy George Kaplan" (la frase inglesa que pronuncia enfáticamente es "Okay, I am George Kaplan now"). A partir de ese momento, el nombre "George Kaplan" va adquiriendo una rigidez semántica cada vez más fuerte. "George Kaplan" pasa a referirse factual y contrafácticamente a cierto individuo particular, y sólo a él. "George Kaplan" pasa a referir rígidamente a la persona que antes se llamaba Roger Thornhill. Mejor dicho, éste personaje sigue llamándose "Roger Thornhill" en determinados contextos. Pero en otros contextos, la denominación rígida "George Kaplan" tiene cada vez más protagonismo

semántico. En esos contextos, acaba adquiriendo la rigidez semántica de un "yo" (como pone de manifiesto la propia frase pronunciada por Cary Grant).

Los aspectos de la película que quiero resaltar son básicamente tres. En primer lugar, diferentes tokens de un mismo tipo de expresión (el nombre propio "George Kaplan") parecen poder tener propiedades referenciales muy distintas en contextos muy cercanos de uso del lenguaje. Incluso un mismo token puede ser interpretado referencialmente de modo muy distinto en contextos muy cercanos. Si un miembro del contraespionable escribe una nota donde aparece el nombre "George Kaplan", y esa nota es inmediatamente leída por uno de los espías, el mismo token referirá de modo muy diferente según es escrito o leído. Quién escribe y quién lee hace que cambie el contexto de uso de un mismo token. Y con ese cambio, cambiará la referencia. A veces las diferencias entre contextos de uso resultan muy claras. Otras veces, no lo son tanto. Incluso cabría imaginar situaciones en las que sean muy borrosas todas las diferencias entre contextos de uso. Como caso límite, podemos pensar en Roger Thornhill siendo un agente doble que no confía plenamente ni en sus compañeros espías, ni tampoco en sus compañeros del servicio de contraespionaje, y que con ese trasfondo llega a asumir que él es George Kaplan (pronunciando la frase "Okay, I am George Kaplan now").

En segundo lugar, atribuir referencia a un tipo de expresión parece que requiere postular la existencia de alguna clase de mecanismo capaz de estabilizar y dar protagonismo a algunos de esos usos referenciales, de ciertos tokens del tipo de expresión en cuestión, sobre los demás. Esta hipótesis resulta ciertamente muy natural y explicativa. Pero habría que tener en cuenta detalles importantes como los siguientes: 1) que tales mecanismos estabilizadores de un uso también han de poder ser muy diversos, desempeñando su función de muchas posibles formas diferentes; 2) que ha de poder activarse un mecanismo u otro en marcos temporales relativamente muy pequeños, siguiendo los cursos variables de la comunicación; 3) que aunque a veces esos mecanismos sí entren en abierto conflicto, también han de poder convivir pacíficamente sin que resulte perturbada la comunicación; y 4) que esos mecanismos han de actuar sobre la manifestación y el reconocimiento de las intenciones referenciales presentes en el uso de los tokens del tipo de expresión en cuestión.

En tercer lugar, incluso para evaluar si un token de un tipo de expresión tiene una referencia determinada parece necesario considerar conjuntamente tanto las intenciones referenciales presentes en la ocasión de uso de ese token como la existencia de mecanismos estabilizadores de un uso. La referencia de un token de un tipo de expresión podría ser diferente de la referencia asignada intencionalmente si existen fuertes mecanismos estabilizadores de un uso contrario a esa referencia

intencional presente en esa ocasión de uso. La referencia intencional no está libre de revisión y reconsideración. Puede llegar a ser vista, por ejemplo, como un simple "error" o "despiste", o incluso como un mero "balbuceo mental", etc. Esto implica que deberíamos poder distinguir, por un lado, la referencia intencional presente en una ocasión de uso de un token de cierto tipo de expresión de, por otro lado, la referencia del token en esa ocasión de uso. Y ambas cosas deberían distinguirse también de la referencia del tipo de expresión al que pertenece el token en cuestión.

En conexión con estos tres aspectos del referir que encontramos magistralmente ejemplificados en la película de Hitchcock, pueden defenderse las tres tesis siguientes:

1. Pluralismo referencial: Distintos tokens de un mismo tipo de expresión pueden perfectamente referir de maneras muy diferentes en aproximadamente los mismos contextos de uso. Y pueden hacerlo sin que necesariamente surjan conflictos que dificulten la comunicación. En el caso de que surjan conflictos, tales conflictos tendrán que resolverse en un sentido u otro. Sin embargo, como muestra la película de Hitchcock, lo habitual es que convivan maneras muy distintas de referir, por parte de distintos conjuntos de tokens de un mismo tipo de expresión en contextos de uso muy cercanos, o incluso en los mismos contextos, sin que surja ningún conflicto serio. Mejor dicho, los conflictos serios que surgen no son precisamente conflictos semánticos. El problema grave no es saber a quién se refiere el nombre propio "George Kaplan", sino saber quién es George Kaplan. El nombre propio simplemente se usa, y se usa de modos muy diversos, no se menciona para formular una pregunta semántica (filosófica) sobre su referencia.

2. Existencia de una dinámica crucial que involucra intenciones referenciales y mecanismos estabilizadores de ciertos usos: Las intenciones referenciales se manifiestan en ciertos usos. Y sobre esos usos actúan determinados mecanismos estabilizadores. Esos mecanismos estabilizadores dan lugar a la referencia de los tipos de expresiones. Incluso las propias intenciones referenciales presentes en cada uso de un token de un cierto tipo de expresión se reconocen, y pueden modificarse y corregirse, en relación a esos mecanismos. Esta dinámica parece el factor crucial a la hora de explicar la referencia atribuible tanto a los distintos tipos de expresiones de un lenguaje como a los distintos tokens, pertenecientes a esos tipos, que son usados en diferentes contextos. Esto permite entender que haya muchos casos en los que ciertamente resultaría muy forzado explicar la rigidez de un tipo de expresión en base a "bautismos iniciales", en los que algo se asocia ostensivamente con dicho tipo de expresión, o en base a una descripción que fije la referencia estándar de ese tipo de expresión, y a "cadenas causales de transmisión de un uso" que vinculen

tokens de tal tipo de expresión manteniendo constante esa referencia inicial. Lo importante para que un tipo de expresión refiera de forma rígida debe ser algo más general. Lo importante es que se manifiesten y se reconozcan de una manera estable ciertas intenciones referenciales de rigidez respecto al uso de algunos conjuntos de sus tokens, aunque no necesariamente respecto a todos ellos. Dicho de otro modo, para que un tipo de expresión refiera rígidamente, deben existir ciertos mecanismos estabilizadores de esa función referencial rígida a través de distintos usos de tokens de la expresión en diferentes contextos[2]. Lo mismo cabría decir de otras funciones referenciales como la referencia ficcional o la referencia descriptiva. Los "bautismos", las "descripciones que fijan una referencia estándar" y las "cadenas causales de transmisión de un uso" ofrecen una clase de mecanismos para la referencia rígida, pero existirían otras muchas clases de mecanismos. Algunos son incluso más habituales que el tematizado por Kripke. En cualquier caso, 1) tales mecanismos han de poder ser muy diversos, 2) también han de poder llegar a ser operativos unos u otros en marcos temporales relativamente pequeños, 3) ha de ser posible que muchas veces no exista ningún conflicto grave entre ellos que impida la comunicación, y 4) han de actuar siempre, como notablemente señala el propio Kripke, sobre la manifestación y el reconocimiento de ciertas intenciones referenciales ligadas al uso de tokens concretos de los tipos relevantes de expresiones.

3. Rigidez sin metafísica: La rigidez de un tipo de expresión hace que esa expresión se refiera típicamente a la misma entidad en todo mundo posible en el que esa entidad exista. Podemos decir así que la rigidez del tipo de expresión "Nixon" hace que sea típicamente imposible que (a) Nixon no sea Nixon. Y que la rigidez de los tipos de expresiones "agua" y "H_2O" hacen que, si en nuestro mundo la composición química del agua es H_2O, como al parecer hemos descubierto, entonces sea típicamente necesario que (b) el agua sea H_2O. En otras palabras, la rigidez hace que el agua sea típicamente, de manera esencial, lo que el agua es en nuestro mundo. La manera cómo nos estamos expresando es deliberadamente crucial. Pues de nuestras tesis 1 y 2 se sigue que ciertos tokens de "Nixon" podrían no designar rígidamente en algunos usos, y que ciertos tokens de "agua" también podrían no designar rígidamente en algunos usos (dejemos ahora al margen el tipo de expresión "H_2O"). Y las preguntas pertinentes son: ¿Podría ser así con alguna de las dos ocurrencias de "Nixon" en algunos tokens concretos de (a)? ¿Y podría ser así con la ocurrencia de "agua" en algunos tokens concretos de (b)? Si así fuera, dejaríamos de tener, en esas ocasiones de uso, imposibilidad en (a) y necesidad en (b). Y a pesar de la rigidez asumida para los tipos de expresión

2 La conexión de estas ideas con otros análisis, como por ejemplo los de Devitt (1981) y Devitt y Sterelny (1987/1999), serían sumamente relevantes. Pero quedan fuera de los propósitos de este trabajo.

"Nixon", "agua" (y "H_2O"), para cualquier acto de habla particular que produzca tokens de (a) y (b), pasaría lo mismo (¡pasaría esto incluso en los propios tokens (a) y (b) de más arriba¡). Como indicábamos antes, los tokens concretos de un tipo de expresión tendrían una referencia determinada conjuntamente tanto por ciertos mecanismos estabilizadores del uso de diferentes tokens de un mismo tipo de expresión, como por las intenciones referenciales presentes en cada ocasión de uso de esos tokens. Pero, de ser así, se siguen consecuencias desastrosas para la necesidad y esencialidad presuntamente involucradas en (a) y (b). Y en general, para la clase de modalidades metafísicas que se pretenden descubrir a través de (a) y (b). Pues, ante cualquier uso concreto de (a) y (b), o bien 1) tendremos únicamente el tipo de "necesidad física" impuesta por la primacía de determinados mecanismos estabilizadores de un uso, o bien 2) tendremos únicamente el tipo de "necesidad lógica" impuesto por nuestras convenciones que estipulen que esos usos concretos son rígidos. Putnam (1990) argumenta justamente esto, marcando distancias con Kripke. Creo que por otros caminos habríamos llegado al mismo resultado.

La tesis 1 sintoniza plenamente con las observaciones sobre los nombres que hace Wittgenstein (1953) en la primera parte de sus Investigaciones Filosóficas. Ésta es la tesis más importante. Las otras dos tesis pueden considerarse consecuencia suya. La tesis 2 permitiría explicar de manera muy natural los peculiares modos de referir presentes en muchos de los casos repetidamente discutidos en la literatura sobre estos temas, casos que involucran términos como "Gödel", "Madagascar", "Jonás", "Ramsés VIII", etc. La tesis 3 moderaría enormemente las pretensiones metafísicas de la semántica, y el estilo de discurso esencialista surgido a propósito de la obra de S. Kripke.

Nuestra discusión se ha focalizado en el comportamiento semántico de los nombres propios (como "George Kaplan"). Creo que podría fácilmente ampliarse a otros tipos de expresiones referenciales singulares como las descripciones definidas, los pronombres personales y los pronombres demostrativos. Asímismo, podría ampliarse al caso de los términos de clase natural y a los términos teóricos (términos que, como "H_2O", por lo general refieren singularmente a entidades abstractas). Por lo que respecta a estos últimos, la referencia ficcional, la referencia descriptiva y la referencia rígida muchas veces conviven perfectamente, sin que se dificulte seriamente la comunicación y sin que ningún mecanismo estabilizador referencial llegue a imponerse definitivamente sobre los demás. Y esto permite entender que exista aquí un amplio espacio abierto a la discusión filosófica.

Las tres tesis más arriba esbozadas son sin duda muy generales. Pero ponen el énfasis en un punto muy concreto (en un punto explícitamente wittgensteiniano). Frente a la complejidad del referir, la rigidez es un fenómeno semántico demasiado simple para ser suficientemene explicativo.

Tratar todos estos problemas desde algo como una película tiene ventajas sobre un tratamiento basado exclusivamente en experimentos mentales, en enunciados contrafácticos o en nuestras intuiciones como hablantes (las tres fuentes de evidencia más habituales para estos temas). Una película como la de Hitchcock nos ofrece en vivo un conjunto plausible de usos lingüísticos y de contextos de uso. Constituye una especie de gran "juego de lenguaje". No es preciso construir mundos, ni evaluar contrafácticos, ni apelar a nuestras intuiciones como habantes. Por decirlo así, las respuestas semánticas están ya ahí. Y lo que es más importante, también el que no haya respuestas. Muchas veces, simplemente, no habrá respuesta. No habrá ninguna respuesta última y contundente a nuestras inquietudes semánticas. Pero eso, como la propia película de Hitchcock muestra, como pone perfectamente de manifiesto, de ningún modo tiene porqué impedir la comunicación.

REFERENCIAS

Devitt , M. 1981, *Designation*, Nueva York, Columbia Univ. Press.
Devitt, M. y K. Sterelny 1987/1999, *Language and Reality: An Introduction to the Philosophy of Language*, Oxford, Blackwell.
Kripke, S. 1972, Naming and Necessity, en G. Harman y D. Davidson (eds.) *Semantics of Natural Language*, Dordrecht, Reidel (2ª ed. revisada y ampliada en Oxford, Blackwell, 1980. Traducción castellana de esta edición: *El nombrar y la necesidad*, México, UNAM, 1985).
Putnam, H. 1990, "Is Water Necessarily H_2O", en H. Putnam, *Realism with a Human Face*, (J. Conant, ed.) Cambridge, Harvard Univ. Press, 1990.
Wittgenstein, L. 1953, *Philosophical Investigations*, Oxford, Blackwell (Traducción castellana: Investigaciones filosóficas, Máxico, UNAM, 1986).

¿ES LA TESIS DE LA CARGA TEÓRICA DE LA OBSERVACIÓN INCOMPATIBLE CON EL DOGMATISMO?

MIREIA LÓPEZ AMO
Universitat de Girona - LOGOS Group

ABSTRACT: The goal of this paper is to show, against Pryor, that the theory-ladenness of observation is incompatible with his dogmatist proposal about perceptual justification. I will present dogmatism through two differentiated thesis, Immediacy (IMM) and Sufficiency (SUF), and I will argue that although the theory-ladenness of observation under certain interpretation is compatible with (IMM), it is not compatible with (SUF). Thus, it cannot be said, as Pryor does, that the theory-landenness of observation is unproblematic for dogmatism.
KEY WORDS: dogmatism, theory-ladenness of observation, justification, cognitive penetration, defeaters.

En su artículo "The Skeptic and the Dogmatist", James Pryor (2000) presenta su proyecto dogmatista contra el escepticismo como una propuesta sobre la justificación perceptiva. Considerando que 'S' está por el sujeto y que 'p' refiere a cualquier proposición perceptiva sobre el mundo externo el dogmatismo sostiene que:

(DOG): Siempre que S tiene una experiencia, E, de como si *p*, E proporciona justificación inmediata (*prima facie*) a S para creer *p*.

Una experiencia de como si *p* nos justifica (*prima facie*) inmediatamente para creer *p* si y solo si la justificación que tenemos para creer *p* en base a esta experiencia no descansa en la justificación que pudiéramos tener para creer otras proposiciones. De acuerdo con (DOG), la justificación que nos proporciona la experiencia es tan solo justificación *prima facie*; es decir, justificación que puede ser socavada a la luz de indicios contrarios. Si tenemos justificación *prima facie* para creer *p*, en ausencia de indicios a nuestro alcance que socaven dicha justificación, tenemos justificación ATC.[1] Así pues, (DOG) afirma que, en ausencia de socavadores, estamos inmediatamente justificados-ATC en creer lo que a través de la experiencia nos parece que es el

[1] La justificación ATC se refiere a lo que se conoce en inglés como justificación "all-things-considered". A falta de una traducción adecuada se ha optado por el uso del acrónimo del término en inglés.

caso.

(DOG) puede descomponerse en dos tesis diferenciadas que no se implican mutuamente: Inmediatez (INM) y Suficiencia (SUF).[2]

(INM): Siempre que un sujeto S tiene una experiencia de como si *p* que proporciona a S justificación *prima facie* para creer *p*, S está justificado inmediatamente en creer *p*.

(SUF): Siempre que un sujeto tiene una experiencia, E, de como si *p*, E proporciona justificación *prima facie* para creer *p*.

Veamos que (INM) no implica (SUF). Un fiabilista podría considerar que no es suficiente la mera presencia de cierta experiencia para que, en ausencia de socavadores, estemos justificados en creer lo que a través de ella nos representamos. Para que cierta experiencia confiera justificación debe ser fiable; es decir, debe pertenecer a un tipo de experiencias tales que la mayor parte de las creencias que nos formamos basándonos en ellas son verdaderas. Presumiendo que no todas las experiencias son fiables, dicho fiabilista juzgaría que no todas poseen poder justificativo. Así pues, concluiría, en contra de (SUF), que no siempre que un sujeto tiene una experiencia de como si *p* este está *prima facie* justificado en creer *p*. Por otro lado, podría seguir sosteniendo que cuando las experiencias son fiables, y no hay socavadores, la justificación que nos proporcionan las experiencias es inmediata.

(SUF) tampoco implica (INM). Supongamos que para estar *prima facie* justificados para creer *p* en base a la experiencia de como si *p* debemos estar justificados en creer otras proposiciones. La justificación *prima facie* que conferiría la experiencia no sería inmediata. No obstante, admitamos que necesariamente estamos justificados por defecto a creer estas otras proposiciones (el tipo de justificación que Wright (2004) reivindica para proposiciones como que hay mundo externo o que no soy un cerebro en una cubeta). Bajo este supuesto, siempre que tenemos una experiencia de como si *p* estamos *prima facie* justificados en creer *p* y, por tanto, (SUF) es verdadero.

Nuestro objetivo será determinar si la tesis de la carga teórica de la observación (CTO) constituye una amenaza para (DOG). Pryor sostiene que no lo es ya que ninguna de las tres interpretaciones aceptables de (CTO) resultan a su juicio incompatibles con (INM). Pryor no discute

[2] Siegel y Silins (manuscrito) ya presentan (DOG) a través de estas dos tesis. Nuestra formulación de ambas tesis supone tan solo una ligera modificación respecto a la que ellos ofrecen.

la compatibilidad de (CTO) con (SUF) por lo que parece asumirlo como una cuestión incontrovertida. No obstante, al menos bajo una de estas interpretaciones, surgen serias dudas acerca de su compatibilidad.

Primero, expondremos de forma general la tesis (CTO). Después, presentaremos una las posibles interpretaciones de (CTO) y analizaremos críticamente las razones de Pryor para defender que es compatible con (INM). Finalmente, mostraremos que esta versión no es compatible con (SUF).

De forma general, podríamos enunciar la tesis de la carga teórica de la observación en los siguientes términos:

(CTO): La observación depende, de alguna forma, de la teoría.[3]

(CTO) fue presentada entre los años 50 y 60 por Hanson (1958) y Kuhn (1962) entre otros, con la sospecha de que (CTO) atentaba contra el presupuesto empirista de que la observación es un método objetivo para evaluar nuestras teorías. Si la observación depende de la teoría, no podemos decir que la observación constituya una base neutral y objetiva para contrastar nuestras teorías. Además, si (CTO) es verdadera, nuestras teorías pueden verse auto-corroboradas, ya que si la observación depende de cierta teoría es posible que tienda a corroborar la misma teoría de la que depende.

Para entender (CTO) debemos dilucidar en qué consiste la relación de dependencia involucrada. Tradicionalmente se ha pensado que la observación depende causalmente de la teoría de tal modo que la segunda determina el contenido de la primera. En la literatura, el uso de "observación" es ambiguo; algunas veces se refiere a experiencia perceptiva y, otras, a creencia perceptiva. Esta ambigüedad da lugar a dos formas de interpretar (CTO). En una primera versión, (CTO) enuncia que las creencias teóricas que sostenemos determinan causalmente el contenido de algunas de nuestras experiencias. En una segunda versión, sería directamente el contenido de algunas de nuestras creencias perceptivas el que resultaría causalmente afectado por nuestras creencias teóricas.

La primera de estas versiones es la que ha suscitado más controversia e interés. Precisamente esta es la interpretación de (CTO) que será objeto de nuestra discusión. Veamos en

[3] La noción de teoría en el ámbito de la filosofía de la ciencia es usada de un modo restrictivo, puesto que se refiere a un conjunto de creencias articuladas que intentan dar cuenta de un fenómeno más o menos complejo. No obstante, de acuerdo con el tipo de discusión que llevaremos a cabo, haremos un uso más laxo de la noción incluyendo bajo su extensión meras creencias individuales, de forma que podamos decir que la observación depende de alguna manera de nuestras creencias teóricas.

qué términos la presenta Pryor:

> La afirmación de que la observación está cargada teóricamente puede querer decir que la teoría que uno sostiene afecta causalmente a las experiencias que tiene. Dos sujetos, dada la misma estimulación sensorial pero con distintas creencias de base, pueden diferir respecto a lo que perceptivamente les parece que es el caso. (Pryor 2000, 540).

En base a estas consideraciones podríamos presentar esta versión de (CTO) del siguiente modo:

(CTO1): La creencias teóricas que sostiene el sujeto determinan el contenido de algunas de sus experiencias.

Resultados empíricos de la neurociencia y la psicología cognitiva parecen avalar (CTO1). En ellos se muestra que el contenido de la experiencia puede resultar afectado por aquellos estados cognitivos del sujeto que constituyen sus antecedentes causales, como podrían ser sus creencias teóricas. Este fenómeno se conoce como "penetración cognitiva". Apelando a la penetración cognitiva, se pretende dar cuenta de cómo es posible que dos sujetos, fijados los estímulos distal y proximal, las condiciones de percepción, el estado de sus órganos sensoriales y su foco de atención, tengan dos experiencias distintas (respecto a su contenido y a su carácter fenoménico). Se considera que sus experiencias son distintas como resultado de la diferencia entre sus estados cognitivos (MacPherson 2011 y Siegel 2011). Por lo tanto, debemos suponer que estos ejercen cierta influencia sobre la experiencia que acaba por determinar su contenido.

El propio Pryor acepta la existencia de indicios empíricos a favor del tipo de dependencia que se enuncia en (CTO1), pero entiende que es compatible con (INM). Veamos las dos razones que aduce para defender tal compatibilidad. En primer lugar, señala que (INM) y (CTO1) versan sobre cuestiones distintas: (CTO1) trata de la formación de los contenidos de la experiencia e (INM) trata de la estructura de la justificación perceptiva. En segundo lugar, ofrece un razonamiento por analogía. El contenido de nuestra experiencia podría estar causalmente determinado por la gafas que llevamos, pero no por ello diríamos que nuestra justificación perceptiva depende en algún sentido de las gafas que llevamos. Análogamente, si el contenido de nuestra experiencia resulta causalmente afectado por nuestras creencias teóricas, nuestra justificación perceptiva tampoco depende en ningún sentido de la justificación que tengamos para estas creencias. Consiguientemente, la mera existencia de esta influencia causal no impide que nuestras creencias perceptivas puedan estar inmediatamente (*prima facie*) justificadas.

Sin embargo, creemos que este razonamiento es defectuoso. La afirmación de que la justificación de nuestras creencias perceptivas no depende de las gafas que llevamos es un sinsentido.[4] Puesto que lo que se supone que se debe concluir en el caso no controvertido de las gafas no puede en verdad establecerse al ser la supuesta conclusión ininteligible, tampoco puede establecerse por analogía la conclusión correspondiente en el caso controvertido. No obstante, se pueden ofrecer razones alternativas a favor de la compatibilidad entre (CTO1) e (INM).

Determinar si (CTO1) es compatible o incompatible con (INM) consiste en ver si (CTO1) implica que no podemos estar inmediatamente (*prima facie*) justificados en creer que lo que nos representamos a través de nuestra experiencia, independientemente de si creemos que el contenido de alguna de nuestras experiencias depende causalmente de nuestras creencias teóricas. Si en algún caso creemos que se da esta dependencia, ello puede socavar nuestra justificación perceptiva inicial. Sin embargo, mientras no tengamos ninguna creencia o sospecha al respecto, aunque de hecho exista tal dependencia, no diremos que nuestra justificación perceptiva depende de la justificación de la creencias teóricas de las que depende causalmente nuestra experiencia. La razón de ello es que estas creencias teóricas no serían accesibles al sujeto, de forma que no podrían actuar como justificadores.

Para concluir que (CTO1) no resulta problemático para (DOG), Pryor debería haber mostrado que (SUF) también es compatible con (CTO1). Creemos que la razón por la cual Pryor no juzga necesario discutir esta cuestión está vinculada con el argumento que ofrece a favor de (SUF). Pryor considera que (SUF) es verdadero en tanto que es una propiedad esencial de la experiencia, su fuerza presentacional, la que da cuenta de su capacidad de conferir justificación. A través de la fuerza presentacional de la experiencia el sujeto toma su contenido como representando verazmente el mundo. Vemos pues que, para Pryor, la ruta etiológica de la experiencia es irrelevante en relación a su capacidad para ofrecer algún tipo de justificación. (CTO1) es simplemente una tesis sobre la formación de la experiencia. Por lo tanto, de acuerdo con Pryor, nada de lo que se derivaría de (CTO1) puede tener consecuencias respecto a la capacidad de la experiencia para conferir justificación.

Sin embargo, (SUF) es objeto de críticas en forma de contraejemplos que apuntan a su extrema permisibilidad (Markie 2005 y Siegel 2011). Estos pretenden mostrar que ciertas rutas etiológicas de la experiencia, algunas de las cuales involucran creencias teóricas, son intuitivamente defectuosas, y que esto tiene consecuencias en relación a la capacidad de la experiencia para ofrecer

[4] Nótese que en el ejemplo que propone Pryor no se hace referencia a proposiciones o creencias sobre nuestras gafas, sino a las gafas mismas.

justificación *prima facie*. Por lo tanto, incluso concediendo que la fuerza presentacional de la experiencia es una propiedad esencial de esta, estos contraejemplos cuestionan que sea en virtud de su fuerza presentacional que la experiencia pueda conferir algún tipo de justificación.

Consideremos el siguiente caso: mi miedo a las serpientes cuando estoy paseando por el bosque me lleva a creer injustificadamente que ante mí hay una serpiente, y esta creencia influencia mi experiencia de forma que acabo viendo como si una serpiente ante mí. Finalmente, en base a esta experiencia me acabo formando la creencia que hay una serpiente ante mí. Nuestras intuiciones respecto a si en este caso estamos justificados en esta última creencia no están claras.

Un primer elemento que nos hace dudar de que estaría perceptivamente justificada en un caso como el anterior es que si aceptáramos que estoy justificada, deberíamos admitir que la justificación perceptiva es excesivamente fácil de obtener. Por el mero hecho de haber tenido una experiencia que no hubiera tenido si no hubiera creído injustificadamente que hay un serpiente ante mí, deberíamos decir que estoy justificada en esta misma creencia para la que inicialmente no tenía ningún tipo de justificación.

Pryor podría ofrecer una explicación para nuestras intuiciones en este supuesto contraejemplo a (SUF) que salvaguardara esta tesis. Podría considerar que el hecho de que dudemos de que en un caso como el descrito pueda tener justificación responde a que no puedo tener justificación ATC si tengo conocimiento del modo defectuoso en que se ha formado mi experiencia. Supongamos que tengo indicios para pensar que mi experiencia depende de mi creencia injustificada. Según Pryor, esto constituiría un socavador para la justificación perceptiva *prima facie* que tengo para creer que hay una serpiente ante mí. Sin embargo, afirmaría que si desconozco cómo se ha formado mi experiencia, en ausencia de otros posibles socavadores, a través de mi experiencia estoy *prima facie* justificada en creer que hay una serpiente ante mí.

En última instancia, esta estrategia dogmatista no resulta exitosa ya que nuestra intuición respecto a mi situación epistémica en un caso como el anterior no concierne fundamentalmente a la justificación ATC, sino a la justificación *prima facie*. En la literatura se distinguen dos tipos de socavadores, pero, como argumentaremos, el conocimiento de la ruta etiológica de mi experiencia no se ajusta a ninguno de los dos tipos. Caben entonces dos opciones: o bien se introduce de un modo *ad hoc* la existencia de un tercer tipo de socavadores o bien se acepta, contra (SUF), que mi experiencia así formada no puede conferir ni siquiera justificación *prima facie* para creer que hay una serpiente ante mí. Sin embargo, ambas opciones parecen abocar al fracaso la estrategia dogmatista para dar cuenta de nuestra intuición y con ello la posibilidad de mostrar que (CTO1) es compatible con (SUF).

El primer tipo de socavadores es lo que se conoce como *rebutting defeaters* (socavadores-R). Estos socavan nuestra justificación para creer *p* en tanto que suponen indicios para pensar que *p* es falsa. Mi experiencia de como si un cuervo blanco socavaría mi justificación para creer que todos los cuervos son negros. El segundo tipo, *undercutting defeaters* (socavadores-U), socavan la relación indiciaria y muestran que los indicios E1, E2, ... que nos han llevado a creer *p* no constituyen, bajo ciertas circunstancias externas, indicios adecuados para creer esta proposición, sin por eso darnos razones para pensar que la proposición es falsa. El hecho de que, si el sujeto tiene conocimiento de ciertas circunstancias externas, E1, E2,... no constituyen buenos indicios para creer *p* no significa que no sea racional en general para el sujeto creer *p* en base a estos indicios. Para poder decir que, bajo ciertas circunstancias externas conocidas por el sujeto, E1, E2,... no constituyen buenos indicios para creer *p* es necesario que sea racional en general para el sujeto creer *p* en base a E1, E2,.... Si no fuera así, ningún indicio del tipo de E1, E2,... podría ser considerado fuente de justificación susceptible de ser socavada. Por ejemplo, saber que la habitación está iluminada por luces rojas socavaría mi justificación perceptiva para creer que la mesa es roja. Sin embargo, esto es así precisamente porque es perfectamente racional en general para el sujeto creer que hay una mesa roja en base a su experiencia de como si hubiera una mesa roja.

Veamos que conocer la ruta etiológica de mi experiencia en el caso que nos ocupa no se asemeja a ninguno de los dos tipos de socavadores que hemos considerado. Saber que mi experiencia de como si una serpiente ante mí depende causalmente de mi creencia injustificada de que hay una serpiente ante mí no me proporciona indicios para pensar que no haya una serpiente ante mí. Por lo tanto, no diríamos que se trata de un socavador-R. Por otra parte, tampoco se trata de un socavador-U. La razón es que, independientemente de si conozco la ruta etiológica de mi experiencia, no parece racional en general creer que hay una serpiente ante mí, en base a mi experiencia de como si una serpiente ante mí, si tenemos en cuenta que esta experiencia depende causalmente de esta misma creencia para la cual en el momento de formación de la experiencia no disponía de indicios a favor de su verdad.

Así pues, si descartamos la posibilidad de caracterizar mi conocimiento de la formación de mi experiencia como un tercer tipo de socavador por tratarse de una estrategia *ad hoc*, deberemos decir que en el caso que nos ocupa, en el que mi experiencia depende causalmente de mi creencia injustificada de que hay una serpiente ante mí, mi experiencia no me confiere justificación *prima facie* para creer lo que a través de ella me represento. Mi experiencia así formada no constituiría en general base racional para creer lo que a través de ella me represento.

En conclusión, hemos visto un caso en el que como consecuencia de la dependencia causal

de la experiencia respecto a cierta creencia teórica diríamos, en contra de (SUF), que a través de esta experiencia no estamos *prima facie* justificados en creer su contenido. Consiguientemente, podemos concluir que (CTO1) es incompatible con (SUF), ya que en algunos casos como el que hemos considerado la experiencia no nos confiere justificación *prima facie* para creer lo que a través de ella nos representamos, como resultado de la dependencia que (CTO1) atribuye a la experiencia respecto a las creencias teóricas.

REFERENCIAS

Hanson, N. R. 1958, *Patterns of Discovery*, Cambridge, Cambridge University Press.
Kuhn, T. 1962, *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.
Macpherson, F. 2011, 'Cognitive Penetration of Colour Experience: Rethinking the Issue in Light of an Indirect Mechanism', *Philosophy and Phenomenological Research*, 84(1), 24-64.
Markie, P. 2005, 'The Mystery of Perceptual Justification', *Philosophical Studies*,126, 347–73.
Pryor, J. 2000, 'The Sceptic and the Dogmatist', *Noûs* 34, 517-549.
Siegel, S. y Silins, N. (manuscrito), 'The Epistemology of Perception'. In Mohan Matthen (ed.), en *Oxford Handbook of Philosophy of Perception*, Oxford, Oxford University Press
Siegel, S. 2011, 'Cognitive Penetrability and Perceptual Justification', *Noûs*, 46 (2), 1-22.
Wright, C. 2004, 'Warrant for Nothing (and Foundations for Free)?', *Aristotelian Society Supplementary*, 78 (1), 167–212.

METAPHOR: MEANING, INTERPRETATION AND CAUSATION*

JANUSZ MACIASZEK
University of Łódź (Poland)

ABSTRACT: In this paper, I examine, extend, and modify Donald Davidson's non-cognitive theory of metaphor in the perspective of his theories of interpretation, action, and rationality. I extend the results to other rhetorical figures, explain the impact of non-literal utterances as a special case of the causal impact of events in terms of Davidson's theory of action, and examine the analogy between presented approach and the theory of metaphor of Samuel Guttenplan.
KEY WORDS: metaphor, allusion, irony, meaning, interpretation, causation, rationalization, event, action.

The main feature of Davidson's (1978) theory of metaphor is the rejection of metaphorical meaning or content. The idea that there are metaphorical meanings seems to result from tacit acceptance of an intuitive notion of meaning. Ajdukiewicz in *Pragmatic Logic* (1965) defined this notion as a *mode of understanding* of an expression, and Wittgenstein in *Philosophical Investigations* (1953) as its *mode of use*. Though both "definitions of meaning" are not equivalent, they seem to describe properly different aspects of the defined notion. From the first-person perspective Ajdukiewicz's proposal seems to be right - *I understand* a metaphorical utterance in another way than a literal one. From the third-person perspective Wittgenstein is right - a *speaker uses* a metaphorical utterance in another way than a literal one. Though both proposals lead to the conclusion that there are metaphorical meanings, they give no hints how to distinguish them form literal ones, the distinction is usually considered as obvious and granted. The philosophers that dealt with metaphor were conscious of this fact and felt obliged to found their theories on strictly formulated theories of meaning, e.g. Searle (1979) and Martinich (1984) referred to the Gricean theory of conversational implication, and ascribed it a multiple role:

(1) the theory of a (conventional) meaning.

(2) the theory of metaphor recognition.

* In this article I continue and modify some ideas conceived in Maciaszek (in print). I am indebted to dr Dorota Rybarkiewicz from my Department who read and corrected the manuscript.

(3) the theory of a metaphorical interpretation (or speaker meaning).

In this paper I adhere to the second role of Grice's theory but, the role of theory of meaning is taken over by Davidson's theory of interpretation, and I reject the notion of metaphorical meaning of metaphor.

1. Interpretation

According to Davidson (1967), while interpreting speaker's utterances, the interpreter ascribes to him beliefs and constructs the theory of meaning of his idiolect, which takes the form of the Tarski's theory of truth. The interpretation is verified constantly during the conversation *via T*-sentences, that enable to compare interpreter's beliefs with the ones ascribed to the speaker[1]. Tarski-style theory of truth ascribes meanings to speaker's words translating speakers idiolect to interpreter's idiolect. The theory is holistic and recursive, what permits explains two facts: mutual interdependence of meanings and their compositionality[2].

The process of interpretation is possible when the speaker and the interpreter have many common beliefs. The beliefs which the interpreter (usually) assumes that he shares with the speaker can be divided into two groups: empirical (occasional) beliefs and analytic beliefs. In case of beliefs about facts that the interlocutors are not witnesses of, the interpreter would rather accept the difference of beliefs than the difference of meanings. Of course Davidson rejected the strict boundary between analytic and synthetic sentences. In consequence, it is not possible to formalize the process of interpretation which is not determined, but on the other hand it constitutes a very flexible tool of explanation of our linguistic practices.

2. Interpretation and maxims of conversation

Paul Gricean in (1975) introduced the notion of conversational maxims that can be violated or flouted, i.e. violated in an apparent way. A metaphorical sentence can be recognized by an apparent violation of the maxim of quality, and allusion as an apparent violation of the maxim of quantity. From the interpreter's point of view the metaphorical sentence is usually a patent falsehood and an allusion as an obvious truth shared by the vast majority of the speakers community, and in consequence as un-informative.

[1] By a belief I consider a *disposition* to accept an utterance of a sentence in certain circumstances - the utterances actualize relevant beliefs.

[2] In fact, Davidson abandons the notion of language as an abstract conventional entity. What we call a language is a class of similar idiolects, where "similarity" means that mutual theories of truth formulated for each other are trivial or nearly trivial.

The same mechanism of the recognition of non-literal utterances can be explained in terms of Davidson's *Principle of Charity*, which assumes rationality of the interlocutors. An interpreter having already ascribed meanings to speaker's words, does not ascribe to him patently false beliefs uttered by him, and does not ascribe to him intention of communicating obvious truths, because he supposes that the rational speaker should also treat him as rational. In this case a sentence is *interpreted as* a metaphorical sentence or an allusion, and the process of interpretation stops. In both cases the speaker apparently behaves as irrational, but the interpreter assumes that he does it for a special *reason*[3].

3. Rationalization of metaphor

The real challenge for Davidson's theory of metaphor is the following: how to explain a very intuitive fact that most metaphors *mean* something? The solution proposed in this paper consists in causal explanation. This approach is compatible with the original Davidson's theory of metaphor and his general philosophical strategy presented in most of his writings[4].

Adopting Davidson's views on language and his metaphysics, the utterance of a non-literal sentence is just an *event* that has its causal power, i.e. it can cause other events. Some of them are typical mental events like offences, complements, etc., but some of them are just the acts of interpreter's belief ascription to the speaker.

One may object to the statement that literal utterances can also offend, complement or amuse. But the difference lies in their communicative role: literal utterances can be *interpreted* and non-literal ones can only be *interpreted as* non-literal, i.e. literal utterances communicate speaker's beliefs in a systematic way described in terms of interpretation, and non-literal ones can communicate beliefs in a non-systematic way. In other terms literal communication can be described as semantic, and non-literal as extra-semantic[5].

The mentioned aspect of extra-semantic communication can be compared to the extra-linguistic one. The extra-linguistic events like behavior, gestures, and faces can cause offences, complements, annoyance or amusement, but they can also play a communicative role. Observing other people's behavior we ascribe to them propositional attitudes and beliefs which they intend to

[3] Davidson's notions of rationality and the *Principle of Charity* have many different aspects. I do not explain them in details in this paper.

[4] In (1978) Davidson states that the only meaning of a metaphor is its literal meaning, and what we call metaphorical meaning or paraphrase is just a description of some consequences of its utterance. As far as I know he never tried to explain the special impact of metaphor in details.

[5] I do not maintain that the semantic way of communicating is not causal, but I maintain that it is realized within a pattern of the theory of truth, i.e. Davidsonian semantics.

communicate. This way of communicating beliefs is causal but, as in the case of non-literal communication, is not systematic and in consequence is not semantic. We can say that there is a strong analogy between extra-linguistic and non-literal communication - the utterance of a non-literal sentence is like an intentional gesture or facial expression.

Some events performed by human agents are called *actions*, if they can be *rationalized*, see (Davidson 1963). Rationalization, which is in fact a special kind of causal explanation, consists in giving a primary reason, i.e. an intention or other attitude to perform action of a desired type, and stating that the performed action is of this type. These redescriptions of events are frequently given in terms of their consequences, e.g. making face by *A* can be described as offending *B*. Of course, it does not mean that every making face by *A* that offended *B* was an action – you can imagine that *A* has neurological problems. But if we rationalize this event, one of the possible rationalizations can take the following form:

- *A* wanted to offend *B*.
- *A* believed that making faces was an offence of *B*.

As there are many other possible rationalizations, e.g. in terms of amusing *B*, etc, there no rules of laws of rationalizing events. This contention is the consequence of Davidson's theses that there are no strict psychological and psycho-physicals causal laws formulated in terms of propositional attitudes. In these terms we can formulate sentences about causal interaction of particular events, but we cannot formulate strict nomological laws. In consequence, we can neither unambiguously rationalize a non-literal utterance, i.e. give its causes, nor predict its all consequences.

Let us imagine that *A* utters something that we interpret as a non-literal utterance. If we treat *A* as a rational being we try to rationalize his verbal behavior, i.e. give reasons of *A*'s utterance. As there are infinitely many possible reasons that we can ascribe to A, the non-literal utterances seem "open", i.e. they can be rationalized in infinitely many ways. Among infinitely many possibilities we can rationalize non-literal utterances in terms of communicative intentions, i.e. intentions of communicating beliefs. If *A* says about a human being *B* that *B* is a donkey, a block of ice, etc., we can state that *A* wanted to offend or criticize *B*, but we can also state that *A* wanted to communicate his belief that *B* is stupid and stubborn or unemotional. In this case the utterance that *B* is a donkey or a block of ice is just the intentional act of communicating speakers beliefs. The content of this belief is just what is traditionally called metaphorical meaning or paraphrase of "donkey" or "block of ice". From this point of view apparent metaphorical meanings can be considered as a by-products of rationalization.

The same can be said about allusions. Sometimes we rationalize the utterances of obvious truths in terms of communicative intentions of the speaker who wanted to communicate his beliefs to us. If a deputy says to the prime minister during the parliamentary debate that two plus two equals four, we can rationalize his behavior in terms of intentions of communicating something. As there are no rules or laws of rationalization, infinitely many possibilities emerge. In the given context some rationalizations are obviously considered as more probable than others, and we usually decide on the most probable one. The conventionalized metaphors and allusions have usually typical and well known rationalizations. Nevertheless all non-literal utterances, even the conventionalized ones, remain "open", i.e. ready for different rationalizations in communicative terms.

Rationalizing is non-literal creative endeavor that depends on interpreters experience, knowledge, and ingenuity - it is rather an art than a science[6]. There are also novel and creative metaphors that are very difficult to rationalize in communicative terms. Nevertheless, they can also be rationalized in infinitely many no-communicative ways, e.g. as an invitation to intellectual effort, or as a mere provocation.

In his monograph *Objects of Metaphor* (2005), Samuel Guttenplan states that while uttering a metaphor in a predicative position in a sentence we predicate of the subject an extra-linguistic object (form the language of the level "zero"). Though this contention does not seem very clear, there is an apparently obvious analogy to the solution presented in this paper. First of all every object or its picture can be described in infinitely many ways and has strong causal power – as Davidson stated in (1978) "A picture is not worth a thousand of words". But it is rather an analogy to the literal use where there is not one interpretation as there is no one description of an object or a picture. The analogy is better if we explain the causal impact of objects and pictures in terms of communicative intentions. This permits us to apply the proposed solution to extra-linguistic metaphors, e.g. in art. If the pattern containing an object, or the picture seems to be presented in too obvious or not compatible way with our everyday experience, where something is out of place, the interpreter of a work of art tends to rationalize the creation of this pattern or a picture. In special cases the rationalization can be formulated in terms of the intention of communicating something that goes beyond any of direct descriptions of an object. If Guttenplan is right, metaphor considered as an extra-linguistic object in a linguistic utterance, obviously it must be considered as out of place

[6] Davidson compares interpreting metaphor to interpreting dreams. This pre-theoretic usage of "interpreting" has nothing to do with the strict notion.

in the pattern, and its use must somehow be rationalized. If the rationalization in communicative terms is impossible, we look for other rationalizations, like artistic provocation.

There is yet one remark to be made. Some metaphorical utterances, as well as some extra-linguistic metaphors and actions, are considered as meaningful, though the agent obviously did not intend to communicate anything. This kind of meaningfulness is very important for psychology, psychotherapy, and theory of art. Of course, the actions of this sort cannot be rationalized in terms of intention of communicating the beliefs. The "hidden" beliefs can be somehow inferred from possible rationalizations formulated in other terms, and the success of this enterprise depends on the ingenuity, creativity, and experience of the interpreter.

4. Metaphor itself

Up till now I have been talking mainly about metaphorical sentences and allusions, though traditionally metaphor is considered as a special word or an expression in a metaphorical sentence. As Searle noticed in (1979) metaphor occurs usually among literal words, e.g. in "Sally is a block of ice", "Sally" is taken literally. It is also easy to see that metaphors can occur in allusions, e.g. "You are not a machine", which is just a negation of a relevant metaphorical sentence "You are a machine". I do not think that there is one criterion of recognizing metaphor as an expression in a non-literal sentence, and I believe there are various strategies to do this basing on the assumption of the rationality of the speaker, i.e. the ability to perform easy logical inferences and the coherence of his holistic net of beliefs[7]. In sentences like "You are a machine" the interpreter recognizes an utterance "machine" as a metaphor because the negation of the sentence is a consequence of commonly accepted beliefs "No man is a machine" and rather obvious empirical belief "You are a man". In general, an interpreter recognizes an expression a as a metaphor in an utterance of a metaphorical sentence $P(a)$ if there is a commonly accepted belief $P(b)$ such that a and b are of the same syntactic category and $\sim P(a)$ is a consequence of $P(b)$ together with other commonly accepted beliefs. The same can be said about an allusion containing metaphor, e.g. "You are not a machine". An interpreter recognizes an expression a as a metaphor in an utterance of an allusive sentence $P(a)$ if there is a commonly accepted belief $\sim P(b)$ such that a and b are of the same syntactic category and $P(a)$ is a consequence of $\sim P(b)$ together with other commonly accepted beliefs[8].

It is important to notice that the existence of a metaphor as a part of a non-literal sentence is not obligatory. The utterance of "Intelligent people do not ask silly questions" is, in certain

[7] As I have mentioned earlier the Davidsonian notion of rationality is rather broad.
[8] This example is taken from Maciaszek (in print).

circumstances, an allusion without metaphor, and an utterance of the oxymoron sentence "I'm for and against it" or "Two plus two equals five" in certain circumstances can be interpreted as metaphorical sentences without metaphor. The process of metaphor recognition is context dependent, and when the context is unknown, rationalization can conduct to different results, e.g. there are two possibilities of rationalization "My sun has already get up from the bed". Depending on the context either "sun" is recognized as a metaphor (e.g. "My wife/husband has already get up from bed"), or "get up from the bed" (e.g. "The sun has already risen").

5. Irony

In Maciaszek (in print) I propose semantics of irony in terms of Davidsonian paratactic analysis, see (Davidson 1968). The idea is that ironical utterance, usually performed with typical ironical tone, can be analyzed as a pair of utterances, where the first one indicates the other. When someone says with ironical tone about a politician "He is a distinguished statesman", it can be considered as a pair of literal utterances that can be interpreted in a normal sense:

- My next utterance does not actualize my belief.
- He is a distinguished politician.

The proposal helps to explain such phenomena as false irony, auto-irony or friendly irony.

The possible objection is that irony has usually very strong causal impact, comparable with the impact of other non-literal utterances. To defend this proposal let us recall that what distinguishes non-literal utterances from literal ones is the rationalization of the former in terms of communication of beliefs, and not in terms of non-communicative beliefs. In case of our example of irony, many non-communicative rationalizations are possible, e.g. the speaker wanted to make fun of the politician, but the communication of beliefs is performed within the process of interpretation, i.e. contrary to metaphor and allusion there is no extra-semantic communication of beliefs. Another possible objection would be that metaphorical sentences do not actualize speaker beliefs exactly like in case of irony. But if we analyzed metaphorical sentences in paratactic terms, the first utterance would be usually trivial, e.g.

- My next utterance does not actualize my belief.
- Sally is a block of ice.

In case of most ironies it is not trivial if the utterance actualizes beliefs of the speaker or not, because they are usually not patent falsehoods. There are, of course, mixed cases of irony and

metaphor like in "Two plus two equals five" uttered in an ironical tone, and the proposed solution helps to analyze even such mixed cases[9].

Closing remarks

The proposals sketched in this article not only shed light on non-literal utterances, but on Davidson's philosophy, as well. In particular they help to explain possible errors, misunderstandings, and ambiguities involved in the highly complicated process of undetermined literal interpretation and undetermined rationalization of utterances in a very complicated endeavor called the language usage in every-day communication and in literature.

REFERENCES

Ajdukiewicz, K. 1965, *Logika pragmatyczna [Pragmatic Logic]*, Warszawa, PWN.
Davidson, D. 1963, 'Actions, Reason, and Causes', *The Journal of Philosophy* 60, 685 – 701.
Davidson, D. 1967, 'Truth and Meaning', *Synthèse* 17, 304 – 23.
Davidson, D. 1968, 'On Saying That', *Synthèse* 19, 130 – 46.
Davidson, D. 1978, 'What Metaphors Mean', *Critical Inquiry* 5, 31 – 47.
Grice, H. P. 1975, 'Logic and Conversation', in W: P. Cole and J. Morgan (eds), *Syntax and Semantics* vol. 3, Academic Press, London.
Guttenplan, S. 2005, *Objects of Metaphor*, Oxford, Clarendon Press.
Maciaszek, J. (in print), 'Niedosłowność i interpretacja' ['Non-literal and interpretation'], to appear in P. Stalmaszczyk (ed.), *Współczesna filozofia języka. Inspiracje i kierunki rozwoju [Modern philosophy of language. Inspirations and trends]*.
Searle, J. 1979, *Metaphor*, in: A. Ortony (ed.) *Metaphor and Thought*, Cambridge University Press, 92 – 123.
Wittgenstein, L. 1953, *Philosophical Investigations*, Oxford, Blackwell Publishing.

[9] One can object even to the importance of ironical tone. In fact, it is only a conventional means of expressing an irony. According to Davidson no conventional means is obligatory, and the process of the interpretation is possible without it. One should also take into account that in the presented approach there is no irony "itself", an utterance can be interpreted as ironical independently of speaker's intentions. The interpreter has never a direct access to speaker's propositional attitudes, he can only ascribe them to him.

ARGUMENTACIÓN RACIONAL SIN TRANSMISIÓN DE LA JUSTIFICACIÓN *

MANUEL PÉREZ OTERO
Departamento de Lógica, Historia y Filosofía de la Ciencia.
Facultad de Filosofía. Universidad de Barcelona / *LOGOS* Research Group in Logic, Language and Cognition
http://www4.ub.edu/grc_logos/manuel-perez-otero
perez.otero@ub.edu

ABSTRACT: Contemporary debates about scepticism, epistemic circularity and warrant transmission takes for granted the following thesis: when warrant transmission fails (so that one cannot acquire, by running an argument with warranted premises, a first warrant for its conclusion, or increase the previous warrant one had for it) the argument fails. I challenge this thesis. The discussion is illustrated with a diagnosis of Moore's Proof of an External World. It can be seen as a good argument, even if warrant doesn't transmit across it. In more general terms: an argument with conclusion C addressed to subject S can be cogent in the sense that the recognition that the premises entail (or make highly likely) C can foster in S the belief in C, without necessarily the warrant for C being gained (or reinforced) by such a recognition.
KEYWORDS: Transmission of warrant, structure of justification, epistemic circularity, argumentative purpose, cogency, scepticism , Wright , Pryor, Jackson, Moore

Los debates contemporáneos sobre argumentos escépticos y anti-escépticos, circularidad epistémica y transmisión de la justificación asumen el siguiente presupuesto:

(W) *Si en un argumento no puede transmitirse la justificación que tengan las premisas a la conclusión, entonces el argumento es fallido.*

El propósito principal de esta comunicación es rechazar ese supuesto. La discusión se ilustrará con algunos argumentos generalmente (pero no unánimemente) considerados fallidos.

Antes de resumir las motivaciones existentes para oponerse a (W), convendrá clarificar algunos de los conceptos que contiene. (Una primera consideración terminológica: usaré 'justificación' en un sentido laxo muy general; coincido con Pryor en que ese uso abarca también lo que otros filósofos querrían denominar alternativamente mediante 'warrant' o 'entitlement'.)

Un argumento puede ser fallido aunque tenga premisas conocidas y se conozca también que de ellas se sigue la conclusión. Por ejemplo, el argumento tal vez sea viciosamente circular en algún

* Financiación: Programa CONSOLIDER-INGENIO 2010, "PERSP - Philosophy of Perspectival Thoughts and Facts" (CSD2009-00056), MICINN (Gobierno de España). Proyecto "Knowledge, Reference and Realism" (FFI2011-29560-C02-01), MICINN. Grupo de investigación consolidado *LOGOS* (2009 SGR 1077), AGAUR (Generalitat de Catalunya).

sentido. Usaré términos como 'bueno', 'satisfactorio', 'legítimo' y otros similares para referirme a argumentos que no son fallidos.

C. Wright (1985, 2002, 2004) ha diagnosticado la aparente circularidad que exhiben ciertos argumentos, postulando que en ellos no se *transmite* la justificación. Esa noción, parcialmente metafórica, se ilustraría a partir de consideraciones como las siguientes. Normalmente, en un buen argumento un sujeto podría adquirir por primera vez justificación para creer la conclusión, o dicha justificación podría ser incrementada, precisamente como resultado de considerar y aceptar el argumento. En ese sentido, cabría concebir la justificación como una magnitud o una propiedad que pudiera *transferirse* o *transmitirse* de ciertas entidades a otras: de las premisas a la conclusión. En un buen argumento, pues, la posibilidad de que la justificación que tienen las premisas se transfiera o transmita hasta la conclusión es esa posibilidad de que la justificación que tiene un sujeto para creer la conclusión pueda adquirirse por primera vez o, al menos, incrementarse gracias a la consideración y aceptación de ese argumento.

Según Wright, la Prueba de Moore de la existencia de un mundo externo es fallida porque no se transmite (no puede transmitirse racionalmente) la justificación:

M1	*Tengo manos*
M2	*Si tengo manos, entonces existe un mundo externo*
	_____
MC	*Existe un mundo externo*

Desde su punto de vista, MC no puede ser apropiadamente justificada a través de M1 y M2 porque para que M1 reciba la justificación que supuestamente recibe (a partir de la experiencia sensorial) es necesario estar previamente justificado en creer MC. Por ello, la Prueba sería fallida.

Explicitemos los tres elementos cruciales de esa explicación:

(Wright-1) *La justificación proporcionada por la experiencia perceptiva es epistémicamente dependiente de la justificación que tengamos de ciertas proposiciones más generales y fundamentales (como por ejemplo MC).*

(Ésta es una tesis sobre lo que podríamos llamar la *estructura* de las relaciones de justificación.)

(Wright-2) *Si la justificación de P es epistémicamente dependiente de la justificación de C, entonces en un argumento con P entre sus premisas y C como conclusión la justificación no puede transmitirse.*

(Ésta es una tesis que vincula la estructura de las relaciones de justificación con lo que podríamos llamar su *dinámica*.)

(W) El supuesto ya mencionado: *sin transmisión de la justificación el argumento falla*.

Pryor (2000, 2004) ha denominado *conservadurismo* (sobre la experiencia perceptiva) a la tesis (Wright-1). Propugna una concepción alternativa, que llama *liberalismo*, vinculada con su teoría *dogmatista* sobre la justificación perceptiva:

(Pryor-1) *La experiencia perceptiva que presenta el mundo como si P confiere una justificación para creer que P que no presupone o descansa en la justificación de ninguna otra cosa que pudiera mencionarse en un argumento en favor de P.*

(Ésta es una tesis sobre la *estructura* de las relaciones de justificación.)

Aplicando esa idea a la Prueba de Moore: la justificación de M1 no dependería de la justificación de MC. Si los posibles fallos en la transmisión de la justificación derivan solamente, de la existencia de relaciones de dependencia epistémica tal y como se describen en el postulado (Wright-2), entonces no veremos motivo (Pryor no ve motivo) para sostener que en la Prueba de Moore no se transmite satisfactoriamente la justificación. Para Pryor (y, naturalmente, para Moore) la Prueba de Moore es legítima.

Tanto Wright como Pryor presuponen sin cuestionarlo el postulado (W). Voy a ofrecer algunas reflexiones sobre la argumentación racional que contradicen ese postulado. Aplicadas al caso de la Prueba de Moore: el argumento podría considerarse legítimo aunque no permitiera la transmisión de la justificación.

Si en un argumento con conclusión C no hay transmisión de la justificación, toda la justificación que tengamos en favor de C tras considerar el argumento será la misma que teníamos previamente. (W) dictamina entonces que el argumento es fallido. ¿Cómo podría eludirse ese resultado negativo?

La respuesta invoca uno de los posibles propósitos o funciones de la argumentación racional, descrito por Jackson (1987), que lo ha bautizado como el propósito *desenredador* [teasing out]. La

idea contenida en (W) (si nuestra justificación para creer cierta conclusión C después de considerar un argumento no es mayor que antes porque las premisas no han aportado ninguna justificación a C, entonces el argumento falla) resulta plausible bajo los presupuestos que nos contemplan como sujetos idealmente racionales. Esos sujetos tendrían acceso consciente e inmediato a todas las consecuencias lógicas de sus creencias. El hecho de que no somos idealmente racionales permite que exista el propósito *desenredador*: al argumentar facilitamos que el destinatario del argumento tenga presente las premisas y tenga presente que de ellas se sigue la conclusión (pues cualquiera de ambas cosas puede haberle pasado desapercibido a un sujeto). Ese destinatario –por ejemplo– puede así reconocer que tiene cierta justificación para creer C (por ejemplo, constatando que las premisas están justificadas y que implican C) sin que para ello sea necesario que esa justificación que ya se tenía para creer C se vea incrementada por *recibir* justificación *procedente* de las premisas. Un argumento como la Prueba de Moore podría entonces ser reivindicado aunque fuéramos neutrales respecto a la controversia entre Wright y Pryor (sobre la estructura y la dinámica de la justificación), contrariamente a la opinión más extendida. El destinatario de la Prueba de Moore puede seguir una estrategia *disyuntiva*: una cierta línea de razonamiento que le permitiría convencerse de que tiene justificación para creer su conclusión independientemente de si es Wright o bien Pryor quien tiene razón.

En esa defensa de la estrategia *disyuntiva* he simplificado las cosas, exponiendo el asunto como si pudiéramos asumir que las posiciones de Wright y de Pryor sobre la estructura y la dinámica de las relaciones de justificación fueran las únicas opciones sobre la mesa. Pero no son las únicas opciones. Por ejemplo, Coliva (2012) defiende una posición intermedia, que denomina *moderadismo*. Si el moderadismo fuera correcto, la estrategia *disyuntiva* quedaría descartada. Sin embargo, hay margen para defender la Prueba de Moore sin comprometerse con dicha estrategia *disyuntiva*.

REFERENCIAS

Coliva, A. 2012, 'Varieties of failure (of warrant transmission—what else?!)', *Synthese*, doi: 10.1007/s11229-011-0006-6 (en prensa).

Jackson, F. 1987, 'Petitio and the Purpose of Arguing', cap. 6 de su libro *Conditionals*, Oxford, Basil Blackwell, 1987. Versión revisada de un artículo publicado en *Pacific Philosophical Quarterly* 65, 1984, 26-36.

Moore, G. E., 1939, 'Proof of an External World', en Moore, G. E. *Philosophical Papers*, London, George Allen & Unwin Ltd., 1959. (Publicado originalmente en 1939.)

Moore, G. E., 1942, 'Reply to my Critics', en Schlipp, P. A. (ed.), *The Philosophy of G. E. Moore*, Northwestern U. P., 533-677.

Pérez Otero, M. 2009, 'Transmisión de la justificación epistémica", en Quesada, D. (coordinador), *Cuestiones de teoría del conocimiento*, Madrid, Tecnos, 148-180.

Pérez Otero, M. 2011, 'Modest Skepticism and Question Begging Proper', *Grazer Philosophische Studien* 83, 9-32.

Pérez Otero, M. 2012a, 'Los propósitos de razonar, ilustrados con el argumento externista anti-escéptico de Putnam', *Theoria* 27, 55-74.

Pérez Otero, M. 2012b, 'La prueba de Putnam contra el escepticismo radical: dos reinterpretaciones basadas en el auto-conocimiento', *Crítica* (en prensa).

Pérez Otero, M. ms-1, 'Purposes of Reasoning and (a New Vindication of) Moore's Proof of an External World".

Pérez Otero, M. ms-2, 'Cogent Argumentation Without Transmission of Warrant".

Pryor, J. 2000, 'The Skeptic and the Dogmatist', *Noûs* 34, 517-549.

Pryor, J. 2004, 'What's wrong with Moore's Argument?', *Philosophical Issues* 14, 349-378.

White, R. 2006, 'Problems for Dogmatism', *Philosophical Studies* 131, 525-557.

Wright, C. 1985, 'Facts and Certainty', *Proceedings of the British Academy* 71, 429-472.

Wright, C. 2002, '(Anti-)sceptics simple and subtle: Moore and McDowell', *Philosophy and Phenomenological Research* 65, 330-348.

Wright, C. 2004, 'Warrant for Nothing (and Foundations for Free)?', *Proceedings of the Aristotelian Society Supplementary Volume* 68, 167-212.

PRESENTISM, TRUTHMAKING, AND LAWS OF NATURE[1]

PABLO RYCHTER
Universitat de València

ABSTRACT: This talk addresses the so called "truthmaker objection" to presentism, and a reply to this objection that I call "the laws-of-nature-proposal" (LOP). This proposal has been recently defended by Ned Markosian, who is one of the major present advocates of presentism. I will argue that, although initially appealing, LOP fails to meet a constraint that is generally taken as part of the challenge behind the truthmaker objection: namely, that the entities invoked to do truthmaker work be *categoric* entities, in a sense to be specified. Thus, I conclude LOP is unsatisfactory as a solution to the truthmaker problem so conceived. However, I will conclude by discussing to which extent the violated constraint is a sensible one and therefore to which extent the truthmaker objection so conceived is really a damaging problem for presentism.
KEY WORDS: Presentism, Truthmaker, Laws of nature

1. Presentism and the Truthmaker Problem.

Presentism is the view that only the present is real. Or, in other words, that only the present exist: there are no merely future objects and no merely past objects. That is to say, the world does not contain past dinosaurs or future outposts in Mars –things that exist only in the past, or only in the future. Presentism contrasts with a variety of non-presentist views according to which at least some of this entities do exist and are as real as present things.

Presentists face the so called "truthmaker problem". This problem may be pinned down as the apparent conflict between presentism and two other, plausible views: the view that truth supervenes on being (that is, the view that if *p* is true, it would be necessary that the world be different from how it is in order for *p* not to be true), and the view that some statements seemingly about the past (and about the future, but let us focus in the past) are true. It is clear what this apparent conflict is. Let us take

(D) Dinosaurs once roamed the Earth

as an example of a true statement seemingly about the past. If presentism is true, it looks as if the truth of D fails to supervene on how the world is. The world of the presentist is "temporally narrow": it does not stretch beyond the present, and contains nothing but present things. And

[1] I want to thank Ned Markosian and the participants of the PERSP Workshop on the Open Future for discussion about the issues in this talk. Research leading to this work has been partially funded by the research projects FFI2010-15717 and CSD2009-00056.

nothing in this temporally narrow world *necessitates* the truth of D. If it were false that there were dinosaurs, everything today could look very much the same. In other words, if presentism is true, it seems that there is a possible world *w* that is indiscernible from the actual world except for the fact that D is not true at *w*. So it seems that the three views (presentism, supervenience of truth on being, and that D is true) are incompatible. (It should be noticed that the Truthmaker problem is sometimes formulated by appeal to principles that are stronger and more controversial than the principle that truth supervenes on how the world is. The differences are interesting, but for our purposes we can stick to the present formulation).

As it is common in this sort of predicaments, one can either take the apparent conflict at face value and avoid contradiction by dropping one of the views in conflict (a move that is in general accompanied by an explanation of why the rejected view *seemed* true at the beginning), or explore the idea that appearances are deceiving and that the three views are not really in conflict. Markosian's proposed solution, the one that we will examine in what follows, has two sides: one of them is of the second, "conciliatory" brand; and the other of the first, "rejectionist" brand. But before moving on to LOP, I would like make one further remark about the truthmaker problem: it is generally assumed by proponents of the problem that one must be *selective* as to what could constitute the supervenience base for the truth of D and the like. Not any feature of the world is allowed. In particular, the presentist cannot attempt to solve the problem by appealing to the property (presently exemplified by so many things) of *being such that dinosaurs once roamed the earth*. These properties are ruled out at the outset because they are, as people say, "past directed", they "point beyond their instances", and are not "categoric". It is in my opinion far from clear what these complaints amount to. (For an elucidation, see Cameron 2010). But in any case, it is clear that the formulation of the truthmaker problem should be understood as including this kind of constraint on any putative supervenience base. Otherwise, the problem is too easily solved. Let me put it in other words: whoever cares about the truthmaker problem (be it a presentist who thinks it can be solved, or a non-presentist who thinks that the problem is fatal for presentism) is driven by the *aspiration to find a categoric supervenience base* for truths like D, a set of properties that do not "point beyond their instances".

2. The laws- of-nature-proposal (LOP)

Let us now turn to LOP, a proposal for solving the truthmaker objection that has been recently defended by Ned Markosian (ms). LOP's basic idea is roughly this: contrary to what we were assuming, many truths seemingly about the past, perhaps D itself, are necessitated by the presentist's temporally narrow world. This world is narrow, but it contains laws of nature. These

laws are either deterministic or indeterministic. If they are deterministic, then they determine, together with everything else in the (present) state of the world, how the world was and will be. So contrary to what we assumed before, there is no possible world w that is indiscernible from the actual world, except for the fact that D is not true in w. In envisaging such a world, we were forgetting about the deterministic laws of nature. If the laws of nature are indeterministic, on the other hand, less things about the past and about the future are determined by how the world is. But according to LOP, it is ok for the presentist to deny the truth of everything that is not so determined. In particular, it is ok to say that D is not true, if it really turns out that the (present) state world is compatible with the falsity of D –i.e. if it is nomologically possible, given the (present) state of the world, that D is not true. This is why I said before that LOP's approach to the truthmaker problem had two sides. LOP says: if the laws of nature are deterministic, then presentism, supervenience, and D are all true, contrary to appearances. If the laws are indeterministic, on the other hand, then D or other statements about the past that we take to be true are in fact not true.

3. Two problems for LOP.

As we have seen, LOP crucially appeals to the laws of nature. Truths about the past and about the future are true in virtue of, among other things, the world (presently) being governed by certain laws of nature. My first worry is about what laws of nature are, and in particular what a presentist can take them to be, if she wants to follow LOP's idea of using these laws in order to solve the truthmaker problem. It seems to me that such a presentist cannot hold the "Humean" view that laws of nature are (or are determined by) regularities among particular events. This view is available for non-presenitsts that embrace past and future events, but it looks as if the the presentist's repertoire of events is too sparse to determine many laws of nature. In particular, I can't see how present events, i.e. all the events that exist according to the presentist, could determine diachronic laws of the sort "if you shake a bunch of Fs at t_1, you *will* get a G by t_2". (You may want to protest here and say: "but that we *will* get a G by t_2 is a present event!" Fair enough, but laws of nature determined by this sort of future-directed events cannot be used in a solution to the truthmaker problem. Remember the constraint that only "categoric" properties be allowed in the supervenience base). That present events cannot determine diachronic laws is especially problematic, because it is precisely this kind of laws that may be thought to ground truths seemingly about the past and the future. So it seems that LOP is incompatible with the Humean conception of laws, and that it *requires* a different conception.

The friend of LOP could try to find such an alternative in the "Armstrong-Dretske-Tooley" account,

according to which laws are relations between universals. (And this is what Markosian actually does). Given this account, LOP would take the following form: the truth of D is at least partially grounded in a relation between universals. But I am not sure this proposal would be accepted by a presentist that is driven by the aspiration to find a "categoric" supervenience base. First, it looks as if the problem of diachronic laws persists. Suppose that we say that R(F, G) is a law of nature relating universals F and G. Suppose further that this law together with the fact that a bunch of Fs are shaken at t_1 jointly determine that we will have a G at t_2. Is the truthmaker problem thereby solved with respect to the truth of "there will be a G at t_2"? I guess it depends on what relation R amounts to. If R is going to be accepted in the "categoric" supervenience base, it cannot be something like the relation of *having instances that will exist right after/before instances of...* . (At this point one may start to think that perhaps the requirement that the supervenience base be "categoric" is not reasonable. Fair enough, but then the truthmaker problem can be solved much earlier, as discussed above). Second, even if diachronic laws are secured as categoric, one may worry about universals themselves. Are universals categoric entities? Should they be allowed to do truthmaking work? These questions may have positive answers in a view like Armstrong's, where universals are multiply located constituents of the concrete world. But I am not sure what the answer should be if we think of universals as abstract entities that may exist uninstantiated. I suspect that such entities will fail to be categoric in the sense intended by the proponent of the truthmaker problem. If so, perhaps Armstrong's peculiar view on universals is another commitment that the friend of LOP must take.

I would like to stress that the problems about laws that I am pointing to are not problems for the presentist as such. They are problems for those presentists that (a) feel pressed to solve the truthmaker problem and (b) want to do it by appealing to laws, as suggested by LOP.

Let me now point to a second worry about LOP. Suppose that the laws of nature are indeterministic enough so that D ("dinosaurs once roamed the Earth") is not true. If we accept the principle of bivalence, we should also say that, in these circumstances, D is false. Now my question is: is D as bad as C?

(C) Centaurs once roamed the Earth.

If the laws of nature are sufficiently indeterministic (something that is the business of science to determine), the presentist who adopts LOP has to say that both D and C are false. Nevertheless, I think it is reasonable for the friend of LOP to distinguish between C and D, and confer D some kind of positive status that C lacks. Perhaps D is in some sense closer to truth than C. Or perhaps D somehow expresses a true proposition, when understood non-literally. In any case, I think it is important for the friend of LOP to have some story to tell about the positive status of D

vis à vis C. Otherwise, I think the proponent of the will be unsatisfied by LOP, since the proposed solution would amount to reject one of the three views that generated the problem (the view that D is true), without any explanation of why it seemed true in the first place.

To summarize, I have mentioned to reasons for being dissatisfied with LOP: first, being a presentist, it is not clear that the friend of LOP can help herself to a conception of the laws of nature according to which laws satisfy the categoricity constraint presupposed by the truthmaker objection. Second, even if this first worry can be met, it looks as if LOP is incomplete as a solution to the truthmaker problem, and that it has to be complemented with an account of the difference between those propositions which are clearly false and those which are false but close to truth. That is to say as a solution to the truthmaker problem, LOP has to be complemented with what we may call the "quasi-truthmaker problem".

3. Rejecting the categoricity constraint?

The first of the two problems discussed above could be solved by rejecting the constraint that the entities in the supervenience base should be *categoric.* I will end by making some very speculative and programatic remarks about this possible move, which I think this is initially appealing.

Perhaps the most direct consequence of dropping the categoricity constraint is that *Lucretianism* becomes an admissible option. As initially presented by John Bigelow, Lucretianism is the view that the world instantiates past-directed properties like the property of having contained dinosaurs. But past-directed properties, now admitted, can also be combined with an ontology of non contingently non-concrete objects, like the one advocated by Timothy Williamson. Given this ontology, it is not the world's instantiating a past-directed property what makes D true but rather the very same things that *were* dinosaurs. These things exist in the present, as non-concrete objects, and each of them instantiates the property of having been a dinosaur. This is on my view an interesting and plausible view that we have independent reason to believe in. (This independent reasons include arguments about the possibility of reference to past objects, and the preservation of the simplest quantified modal logic).

As we can see, the rejection of the categoricity constraint would allow friends of LOP to develop their view, since the apparent non-categoricity of the laws of nature is no longer a problem. But the move will also open the door to competing views, like the varieties of Lucretianism just described. The upcoming task for the friend of LOP is then to explain why her view should be preferred over these other natural and independently motivated alternatives.

REFERENCES

Cameron, R. 2010, 'Truthmaking for presentists', *Oxford Studies in Metaphysics,* 6.

Markosian (ms.). “A Branching-Time Solution to the Truthmaker Problem for Presentism”. Talk delivered at the PERSP Workshop on the Open Future, Barcelona, December 2011.

CREENCIA, CONOCIMIENTO Y VERDAD: EL ENFOQUE LÓGICO-SEMÁNTICO DE FREGE Y RUSSELL Y EL PRAGMATISTA DE PEIRCE Y JAMES

Mª UXÍA RIVAS MONROY
Dpto. de Lógica y Filosofía Moral
Universidad de Santiago de Compostela
uxia.rivas@usc.es

ABSTRACT: This paper describes two different approaches to belief as a cognitive attitude: on one hand Frege and Russell, who highlighted the logical and semantic perspective of belief, and on the other Peirce and James, who understood belief as a rule of action. The paper also intends to emphasize the different philosophical backgrounds of both conceptions of belief, mostly related with their diverse notions of truth and reality.
KEYWORDS: Belief, Knowledge, Truth, Facts, Fregean Thoughts, Judgement, Assertion, Propositional Attitudes, Doubt, Reality, Pragmatism, Frege, Russell, Peirce, James.

1. Introducción

Este texto pretende poner de relieve las diferencias existentes en el análisis de la creencia, a partir de las reflexiones de algunos de los iniciadores de dos corrientes de pensamiento, surgidas prácticamente en la misma época aunque en distintos continentes, a saber, la Filosofía Analítica y el Pragmatismo. Los autores elegidos son Frege y Russell, como representantes de la primera, y Peirce y James, de la segunda, aunque todos ellos distinguen entre la creencia como actitud cognitiva y su contenido, el énfasis del análisis recae en aspectos diferentes. La elección de Frege y Russell se basa en que ambos destacaron especialmente la perspectiva lógico-semántica, mientras que Peirce y James subrayaron explícitamente el carácter de la creencia como actitud cognitiva que se opone a la duda, en el sentido de que si se alcanza una creencia firme, inatacable por la duda, esta llega a funcionar como una verdad, aunque falible y revisable, al tiempo que insistieron en la conexión de la creencia con la acción. En el trasfondo filosófico de ambos análisis de la creencia se puede observar igualmente una concepción diferente de la ontología y de la verdad. Tanto en Frege como en Russell hay una ontología de "hechos", independientes de los sujetos, que para Frege funcionan como la garantía del conocimiento y para Russell son los que hacen verdadera o falsa a la creencia; esta ontología se completa con una concepción de la verdad ya hecha, eterna e independiente de los sujetos cognoscitivos, lo que Putnam llamaría "la verdad con mayúsculas". En el pragmatismo de Peirce, y aún con más claridad en el de James, la creencia no se apoya en hechos ajenos e independientes por completo de los seres humanos, sino más bien en una realidad plástica que

nosotros y nuestras creencias también contribuimos a crear, de tal forma que "el rastro de la serpiente humana" se extiende tanto a la verdad como a la realidad.

2. El enfoque lógico-semántico de Frege y Russell

Frege mantiene que en los contextos de creencia las dos relaciones semánticas básicas de las expresiones, a saber, su sentido y su referencia, difieren de las que éstas tienen en el discurso ordinario, de ahí su denominación de "discurso indirecto". En el discurso indirecto la relación semántica básica de las expresiones, esto es, su "Bedeutung" o referencia, es el *sentido* que estas tienen en el discurso ordinario; en el caso de las oraciones en estilo indirecto su referencia es el pensamiento en tanto sentido objetivo del enunciado. Por ello, de lo que trata el discurso indirecto es de la relación que un sujeto presenta con respecto a un pensamiento: la de creerlo, imaginarlo, dudarlo, etc., pasando, entonces, a un primer plano el pensamiento, en tanto creído, imaginado o dudado, y no tanto su valor de verdad. Otra razón semántica que apoya que la referencia de una oración en un contexto indirecto sea un pensamiento es que su sentido no es completo, y por ello la oración subordinada no puede tener un valor de verdad como referencia, a diferencia de lo que sucede en el discurso ordinario en el que sí expresa un sentido completo o pensamiento, y tiene como referencia un valor de verdad.

Para Frege es importante la dimensión ontológica que adquiere el pensamiento, el cual está vinculado a un tercer reino que se vio obligado a reconocer, ya que ni el mundo físico de los objetos espacio-temporales ni el mundo psíquico de las emociones, sensaciones o representaciones subjetivas podía acoger a los pensamientos, dadas sus características. Este tercer reino incluye, entre otras entidades objetivas, pensamientos, tanto verdaderos como falsos, pero los pensamientos verdaderos son denominados por Frege "hechos". Estos hechos son *descubiertos* y la verdad de los pensamientos es *reconocida* por los seres humanos que los aprehenden. "Descubrir" y "reconocer" son, pues, expresiones indicativas de la posición de los seres humanos en cuanto conocedores, que se encuentran con hechos que existen con anterioridad a su descubrimiento. Para Frege, los pensamientos verdaderos, eternos, inmutables e independientes de los seres humanos constituían la auténtica garantía del conocimiento frente al peligro de solipsismo a que nos llevaría la representación subjetiva como fuente del conocimiento, la cual no va más allá de la mera inmediatez del dato sensible.

La semántica de Frege guió los sucesivos análisis de los discursos indirectos, y, en especial, el de Russell, que, no obstante, lo llevó a cabo a partir de un marco ontológico diferente, a saber, el

del atomismo lógico[1]. Posteriormente Russell denominó a los discursos indirectos "actitudes proposicionales" (cfr. Russell 1983, 70), introduciendo explícitamente la dimensión psicológica y subjetiva propia de este tipo de contextos, pero excluyendo paradójicamente el eco psicologista que evocaba el término fregeano de "pensamiento" al sustituirlo por el de "proposición".

A Russell, que no reconocía la existencia del tercer reino de Frege, le parecía que el hecho de que hubiera pensamientos falsos como elementos constitutivos del mundo atentaba contra su "instinto de realidad" (cfr. Russell 1981, 312-313), ya que la realidad incluía para él todo lo que habría de ser mencionado en una descripción completa del mundo, incluidas, por lo tanto las creencias –verdaderas o falsas–, pero no las proposiciones. En la época de "La filosofía del atomismo lógico" (1918), Russell rechazaba el análisis de la creencia en términos de una relación diádica, que se establecía entre un sujeto y una proposición, para pasar a considerarla como una relación poliádica entre un sujeto y los objetos de la creencia.

El análisis lógico de la creencia como relación diádica resultaba problemático para Russell, ya que la proposición creída se refería a un objeto singular, con lo cual la forma lógica de la creencia sería siempre la misma en todos los casos, y obligaría a reconocer la existencia de proposiciones verdaderas o falsas como ingredientes del mundo, cosa que él no admitía. En el análisis de la creencia como relación poliádica, la relación se establece entre un sujeto y los objetos de la creencia. De este modo si en el mundo se encuentra el hecho referido por la proposición creída, en el sentido de que los elementos constitutivos de la misma están ordenados de igual manera que los del hecho del mundo, la creencia del sujeto relativa a esa proposición será verdadera, en caso contrario falsa. De esta forma, la creencia poseerá diferentes formas lógicas en relación con la naturaleza de lo creído.

La ontología que está de fondo en este análisis de Russell es que en el mundo hay hechos, que son lo que son independientemente de los seres humanos; de esta forma, los hechos se perciben pero no se creen, pues, a diferencia de las creencias, los hechos no pueden ser erróneos; además, la verdad es una propiedad de las proposiciones en tanto describen fielmente los hechos que hay en el mundo. Sin embargo, Russell tuvo que enfrentarse al problema de que la creencia no podía ser un hecho atómico, dados los elementos en ella involucrados, pero tampoco podía ser un hecho

[1] Russell cambió en múltiples ocasiones su análisis de la creencia, al tiempo que cambiaba también su noción de proposición. Selecciono la concepción de la creencia de la época de las conferencias del "Atomismo lógico" (1918), porque son en las que de modo más claro el análisis lógico orienta la reflexión ontológica. Poco después, en "Belief" (1921) (cfr. Russell 1921, 231-252) presenta un análisis de la creencia en términos psicologistas, con múltiples referencias a las obras de James *Principles of Psychology* y *Essays in Radical Empiricism*, que incluye tres clases de creencias en tanto contenido del acto de creer: memoria, expectación y simple asentimiento.

molecular, pues para Russell estos no existían, únicamente las proposiciones podían ser moleculares. Por ello, Russell se vio obligado a ampliar la ontología y a admitir hechos de creencia.

En resumen, tanto Frege como Russell necesitaron ampliar la ontología. Frege para dar cabida a los hechos, que no contaban como entidades del mundo físico, necesitó ampliar los mundos; y Russell para dar cabida a los hechos de creencia, necesitó ampliar las entidades que formaban el mundo, el único mundo que él reconocía.

En esta aproximación a la creencia se encuentra no solo una ontología de "hechos", independientes de los sujetos, sino también una concepción de la verdad ya dada, eterna y que se descubre. Para Frege la verdad es algo así como una propiedad –aunque no exactamente– *sui generis* de algunos pensamientos, independientemente del acceso cognitivo a los mismos. Así pues, la creencia es una actitud de los seres humanos con respecto a pensamientos, los cuales, al margen de esa actitud del sujeto hacia ellos, ya tienen comprometido su valor de verdad. Por ello, la creencia puede ser errónea o verdadera, y es uno u otro caso, dependiendo de si lo que se cree es o no un hecho, esto es un pensamiento verdadero –según Frege– o si se corresponde con un hecho o no –según Russell. Para Frege, los seres humanos descubren la verdad cuando captan pensamientos y les reconocen la propiedad de ser verdaderos, propiedad que ya poseían antes de su reconocimiento. El reconocimiento de la verdad de un pensamiento es un juicio y cuando esto último sucede se expresa conocimiento bajo la forma de una aseveración. En la creencia, sin embargo, se expresa la actitud cognitiva que tiene un sujeto con respecto a un pensamiento, que posteriormente se valorará o evaluará con respecto al juicio que reconoce la verdad de ese pensamiento creído. Para Russell, el contenido de la creencia también es verdadero o falso con respecto a los hechos que acontecen en el mundo. Desde esta perspectiva, la diferencia que existe entre Frege y Russell es que para Russell los hechos son elementos constitutivos del mundo, pero no de un tercer mundo, y la verdad es entendida en términos de correspondencia de las proposiciones con los hechos del mundo.

2. El enfoque pragmatista de Peirce y James

La discusión de la creencia en el pragmatismo americano se encuentra fundamentalmente en lo que se conocen como las actas fundacionales del mismo, a saber, "La fijación de la creencia" (1877) y "Cómo esclarecer nuestras ideas" (1878), donde Peirce introduce el marco general de la doctrina de la duda-creencia.

La idea más relevante de esta doctrina es que una creencia firmemente asentada, que nada va en su contra y que todo está a su favor apoyándola, es indistinguible del auténtico conocimiento. El

análisis de la creencia que ofrece el pragmatismo se centra, por un lado, en el proceso que permite fijar una creencia como firmemente establecida, y por otro, en el valor de la creencia como regla de acción. En el pragmatismo de Peirce la creencia es una actitud cognitiva que se opone a la duda, y que en términos fregeanos vendría a equivaler al juicio y su manifestación en una aserción, lo que más tarde Dewey denominaría una "aserción garantizada". Esto hace patente la proximidad que se puede establecer entre creencia y conocimiento en el enfoque pragmatista, entendido el conocimiento como "creencia verdadera justificada" según la definición analítica usual, pero la introducción del término "verdadera" es, precisamente, el que marca la diferencia entre estas tradiciones, pues se apoyan en una concepción de la verdad muy distinta. Para Peirce y James, la creencia no es creencia en una proposición verdadera, sino que la creencia equivale a la proposición verdadera, en el sentido de que es una idea firme y asentada, de la que estamos seguros hasta el punto de que guía nuestra conducta, es decir, es una idea que está suficientemente probada a través de su puesta en acción, y por eso hasta el momento nada nos ha hecho ponerla en cuestión o dudar de ella; por lo tanto, es una idea de la que estamos satisfechos y conforme a la cual actuamos.

En los escritos de Peirce se pueden distinguir dos maneras de entender la creencia, no siempre claramente separadas. En primer lugar, la creencia se entiende como la actitud cognitiva que equivale al saber o conocer, concepción que Hookway denomina "creencia científica" (cfr. Hookway 2000, 245). En segundo lugar, la creencia es una actitud cognitiva dirigida al establecimiento de un hábito o una regla para actuar, y Hookway la denomina "creencia práctica". Esta vinculación de creencia y acción es la más usual y es la que proporciona a la creencia su rasgo distintivo y esencial en comparación con otras actitudes cognitivas, pues destaca su carácter disposicional y subraya que la creencia va más allá del mero o simple conocimiento, porque está vinculada a la acción, a la puesta en práctica de las ideas, al carácter activo de las ideas, al dirigir nuestra conducta y nuestras actuaciones.

En la base de la concepción de la creencia de Peirce y de James se encuentra un horizonte conceptual que maneja una idea de verdad y de realidad en la que los seres humanos tienen una parte activa, en la que descubren tanto como crean una y otra. La manera cómo se interrelacionan estos elementos "humanos" en la concepción de la realidad y la verdad es diferente en Peirce y en James. En Peirce, su concepción de la realidad y de la verdad está mediada por la idea del acuerdo de la comunidad de investigadores y por la teoría semiótica o teoría de la representación (cfr. Peirce, CP 5.407). En James la realidad está conformada principalmente por tres tipos de entidades, a saber, (i) hechos concretos, (ii) géneros abstractos y relaciones entre ideas, y (iii) el cuerpo de verdades previas que ya poseemos. Pero todos estos elementos contienen ya elementos humanos, imposibles

de eliminar. Por otra parte, la verdad en James es un proceso de adecuación y de orientación, de guía y de conducción satisfactoria de una experiencia a otra, de veri-ficación, validación y asimilación, de verdades en plural.

Por lo tanto, la verdad pragmatista es definible en términos de creencia y de experiencia, en lugar de recurrir a una noción de verdad que finalmente nos compromete con entidades independientes de nosotros. Por eso, Peirce propone que en lugar de emplear el término "Verdad" con mayúsculas, utilicemos como su equivalente el de la obtención de un estado de creencia inatacable por la duda (cfr. Peirce CP 5.416). Por eso, el método que mejor garantiza la fijación de la creencia es el método de la ciencia, y la única garantía de la verdad son las investigaciones y los estudios llevados a cabo por la comunidad de investigadores, que aplicando el método científico obtienen como resultado el establecimiento de creencias. Estas creencias, en la práctica, funcionan como verdades, ya que una creencia para la que hasta ahora no hay ninguna prueba en su contra funciona a todos los efectos como una verdad indudable, pero no infalible, ya que finalmente pudiera darse la circunstancia de tener dudas acerca de ella, y encontrar aquello que nos haga rechazarla. ¿Por qué la creencia funciona como una verdad para Peirce? Porque la creencia está firmemente asentada y nada de lo que acontece va en su contra, con lo cual cada vez el sentimiento de satisfacción aumenta; ninguno de los acontecimientos externos hace que dudemos de ella, por lo tanto, a efectos prácticos funciona como una verdad; y funciona como tal justamente hasta que algo nos la hace poner en duda. Por ello, la creencia no es una verdad, aunque funcione como tal, porque es corregible y falible; pero si nada de lo que sucede nos hace nunca dudar de ella, entonces, para Peirce, tal creencia es indiferenciable de una verdad en el sentido tradicional.

Otra cuestión importante es que en el pragmatismo de Peirce, y aún con más claridad en el de James, la garantía del conocimiento no son exclusivamente hechos ajenos e independientes por completo de los seres humanos, sino más bien que nosotros, con nuestros propósitos e intereses, también contribuimos a crear tanto la verdad como la realidad, pues ambas instancias no son ajenas ni completamente independientes de nosotros. Si bien la realidad tiene un elemento que nosotros no creamos, que está ahí en su inmediatez y resistencia, esta no es la auténtica realidad, porque en ella ningún ingrediente humano está presente. Y lo mismo sucede con la noción de verdad, la verdad no está dada y determinada de antemano de manera eterna e inmutable, y por completo al margen de nosotros. En el pragmatismo la verdad se va haciendo a través de procesos continuos de verificación que confirman el funcionamiento en la práctica de las ideas; aquellas ideas que se confirman, que se validan, que se veri-fican, como afirmaba James, son las que cuentan como ideas verdaderas. Nada

más alejado de la verdad independiente que se descubre o que se reconoce, y que pre-existe lleguemos a ella o no.

REFERENCIAS

Burge, T. 2005, *Truth, Thought, Reason. Essays on Frege*. Oxford, Oxford University Press.
Frege, G. 1973, *Estudios sobre semántica*, Barcelona, Ariel.
--- 1984, *Investigaciones lógicas*, Madrid, Tecnos.
Hookway, C. 1985, *Peirce*, London, Routledge & Keagan Paul.
--- 2000, *Truth, Rationality and Pragmatism. Themes from Peirce*, Oxford, Clarendon Press.
James, W. 1975, *Pragmatism and the Meaning of Truth*, Cambridge, Mass., Harvard University Press.
Ketner, K. L. (ed.) 1995, *Peirce and Contemporary Thought*, New York, Fordham University Press.
Peirce, C.S. 1978-79-80, *Collected Papers*, Cambridge, Mass., Harvard University Press.
--- 1988, *El hombre, un signo*, Barcelona, Crítica.
Putnam, H. 1999, *El pragmatismo*, Barcelona, Gedisa.
Russell, B. 1971, *The Analysis of Mind*, London, George Allen & Unwin Ltd.
--- 1992, *Los problemas de la filosofía*, Barcelona, Labor.
--- 1981, *Lógica y conocimiento*, Madrid, Taurus.
--- 1983, *Significado y verdad*, Barcelona, Ariel.
--- "La filosofía del atomismo lógico" y "Sobre las proposiciones: qué son y qué significan", en Russell, B., *Lógica y conocimiento*, Madrid, Taurus, 245-449.
Sayre, K.M. 1997, *Belief and Knowledge. Mapping the Cognitive Landscape*, Lanham, Maryland, Rowman & Littlefield Publishers.

¿QUÉ HAY DETRÁS DE UN ARGUMENTO? LA DISTINCIÓN OBJETIVIDAD-SUBJETIVIDAD MOSTRADA A TRAVÉS DE LA DISTINCIÓN ENTRE DESACUERDO PERDIDO Y DESACUERDO LEGÍTIMO

MARÍA ESPERANZA RODRÍGUEZ ZARAGOZA
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México

ABSTRACT: The delimitation of objective and subjective can be traced by studying the process of determining truth of statements that give rise to disagreements. This process consists of two levels: the first, corresponding to the language; the second, a non-linguistic level, in which we see how things are either in accordance with reality, the world,, or either with how things are for the subject. What makes these statements objective or subjective, is the way this process takes place.
KEY WORDS: Disagreement, objective, subjective, character, content.

Todos hemos sido partícipes de algún desacuerdo, pero ¿cómo podemos saber si el desacuerdo es tal ? Es decir, ¿ podríamos saber si realmente hubo un desacuerdo legítimo entre las partes disputantes ? Intuitivamente, podemos responder que sí, si lo que sostiene una de las partes, representado mediante un enunciado, se *opone directamente* a lo que sostiene la otra parte y viceversa. No obstante, ¿ qué es esto que *se opone* ? seguramente, no es el mero enunciado. En este escrito, indagaré sobre la naturaleza de los desacuerdos, para que, mediante este análisis, podamos caracterizar con claridad cuándo sí se tiene un desacuerdo legítimo y cuándo el desacuerdo se pierde; y, a través de ello dar luz sobre el problema de la distinción objetividad-subjetividad.

Para lograr esto, se desarrolla un modelo de determinación de verdad, para explicar el proceso que se lleva a cabo para determinar la verdad de aquello sobre lo que se está en desacuerdo. El modelo será probado en enunciados que contienen aspectos del mundo, en principio independientes al sujeto, y en enunciados que contienen cuestiones de gusto personal. Una vez establecido esto, se hará el análisis para distinguir entre desacuerdos legítimos y desacuerdos perdidos.

Aún cuando el hilo conductor de la investigación es poder dar una explicación a favor de esta distinción, paralelamente se trata el problema de dar una delimitación clara entro lo subjetivo y lo objetivo. Problema que se aluzará si resolvemos la distinción entre desacuerdos. Es por ello, que me interesa analizar los desacuerdos que se dan con base en factores subjetivos, como las cuestiones de gusto.

Ahora bien, para el desarrollo del escrito tomo como base las propuestas de Wright (1986), McFarlane (2007), Kaplan (1989) y Lasersohn (2005).

1. Desacuerdo legítimo y desacuerdo perdido

El asunto de distinguir entre un desacuerdo legítimo y uno que no lo es, tiene como trasfondo la pregunta por la objetividad. Wright pretende[1] caracterizar la distinción entre *enunciados genuinos,* enunciados declarativos cuyo valor de verdad (verdadero o falso) es conferido por las propiedades de un *tópico* -subject matter- real, idóneos para expresar conocimiento genuino; y el resto de los enunciados declarativos que tienen la sintaxis de enunciados genuinos, pero que no remiten a cómo las cosas son respecto a un hecho dado. El interés principal de Wright es poder quitar cualquier tipo de contaminación subjetiva de los enunciados. Los elementos que deben darse para que pueda existir un desacuerdo son: (i) partes disputantes, (ii) argumento en disputa, (iii) estados de información.

Como se verá, para Wright, los enunciados que dan pie a desacuerdos legítimos (elemento ii del desacuerdo) expresan cómo son las cosas en realidad, es decir, cómo es el mundo independientemente del sujeto. Pues bien, la distinción entre un desacuerdo legítimo y un desacuerdo perdido, se hace mediante el análisis del argumento que da pie al desacuerdo (ii), y, para llevarlo a cabo, tenemos que analizar aquello que hace que tal argumento pueda sostenerse, esto es, los estados de información (iii) que sostienen a (ii). Para caracterizar esta distinción me basaré en un modelo de determinación de verdad que servirá para analizar los enunciados en los que se basan los desacuerdos.

2. Desacuerdo, modelo de determinación de verdad y objetividad

Primero, me serviré de un desacuerdo que se dé con base en un enunciado que incluya aspectos acerca del mundo actual. Supongamos que surge el siguiente desacuerdo entre los sujetos *T* y *W,* acerca de la altura de la Torre Eiffel:

T: *(pa*) "La altura de la Torre Eiffel es de 324 metros."

W: (-pa) "No, no lo es."[2]

Siguiendo la teoría de Wright, el enunciado asertivo *pa* será objetivamente verdadero si y sólo si, el

[1] En el tercer apartado de <<Inventing Logical Necessity>>, en J. Butterfield ed., *Language, Mind and Logic*, Cambridge University Press, 1986, p. 195-202.

[2] Lo que es objetivo en este ejemplo es la altura de la torre Eiffel en sí misma y no la medida de 324 metros. Esto es, lo objetivo es la altura de la torre, mida ésta lo que mida, con la unidad de medida que sea. Este ejemplo se parece al que Saul Kripke cita en *El Nombrar y la Necesidad*, sobre el metro de la barra de París. Ver Kripke Saul, 2005, *El Nombrar y la Necesidad*, Universidad Nacional Autónoma de México, IIF, México, p.56-59.

significado (lingüístico) de las palabras que expresan la proposición *pa* (de pa), y los *hechos no-lingüísticos* garantizan por sí mismos que ésta es verdadera. Esto es, que corresponda lo expresado en *pa* con cómo es el mundo.

Ahora bien, el significado de un enunciado asertivo está determinado, cuando se cumple:

> a) que los sujetos involucrados en una conversación (hablantes competentes del lenguaje en situaciones normales) saben lo que significan las palabras involucradas, saben gramática y saben cómo se combinan los significados de las palabras. Esto es, tienen el significado lingüístico de las expresiones involucradas en la conversación. Además, ven dichas expresiones dentro de un contexto determinado.

La condición (a) nos dice que los sujetos involucrados saben lo que significan las palabras que componen sus enunciados.[3] Cuando los sujetos conocen el significado de un enunciado, lo que hacen es, primero, fijar un tipo de significado que es dado por reglas semánticas (convenciones lingüísticas), que tienen que ver con cómo los hablantes usan los términos ingredientes del enunciado –a esto se le denomina contexto de uso-. En Kaplan, el contexto de uso, se explica a través de pares ordenados de mundos posibles y tiempos, i.e. <w,t>. Vemos, pues, que esto tiene que ver con lo conocido por los hablantes competentes del lenguaje. A este tipo de significado se le denomina *carácter*. El *carácter* de un enunciado, al ser un significado fijado en un contexto de uso posible, es visto como una función que va de contextos a *contenidos*. Ahora bien, es a través del carácter como se fija este segundo tipo de significado: *contenido*; ya que el carácter da como resultado una intensión. El *contenido* de un enunciado, en un contexto dado, se toma como la proposición expresada en tal enunciado, que se constituye lógicamente. El *contenido* de un enunciado también es visto como una función que va de circunstancias de evaluación a extensiones. Las circunstancias de evaluación son las situaciones contrafácticas (mundos posibles), en las que se evalúa lo dicho en el enunciado; por su parte, las extensiones son los valores de verdad (V o F) que son asignados a tales *contenidos*. En el caso de que *pa* sea objetivo, la proposición expresada tendrá que ver con cómo son las cosas en realidad, esto es, el contenido proposicional apuntará a hechos. Así para "ver" (en sentido amplio) hacia qué hechos apunta la proposición expresada en el enunciado, nos valemos tanto del *carácter* como del *contenido* del mismo, ya que con el primero determinamos qué significa (en un plano lingüístico), mientras que con el segundo fijamos los objetos relevantes (hechos no lingüísticos en un contexto determinado) a los que se refiere y en los

[3] La explicación que doy a continuación conserva la distinción entre los tipos de significado *carácter* y *contenido* introducida por D. Kaplan (1989).

que se basará la evaluación para la asignación de un valor de verdad.

Luego, para que la proposición expresada en *pa* sea *verdadera,* y el proceso de determinación de verdad se complete, debe cumplirse lo siguiente:

> b) la verdad de las proposiciones expresadas en enunciados estará determinada por la correspondencia con los hechos no lingüísticos que dicha afirmación indica.

Así, vemos cómo lo expresado por los enunciados objetivos, *contenido*, será verdadero si y sólo si es el caso que en el mundo (campo de los hechos) ocurre lo expresado por el enunciado; en cambio, si no es el caso que esto ocurra, entonces lo expresado en el enunciado, *contenido*, será falso. Vemos también que la *correspondencia* con los hechos, el mundo, no depende de ningún punto de vista particular o intervención por parte de los sujetos. Por tanto, podemos ver que el significado y verdad de (*pa*) se determinan como sigue:

Pa
La altura de la Torre Eiffel es de 324 metros.
↓

Establecimiento de las condiciones de verdad
Carácter $_{(f)}$: Contexto de uso $<w_{(a)}, t_{1(\text{proferimiento de } pa)}> \rightarrow$
contenido
Significado de *pa* + contexto= intensión
Proposición *pa*
Contenido $_{(f)}$
↓
Torre Eiffel (objeto relevante)
Altura de la Torre Eiffel (***Hecho en el mundo)***
Intensión (lo dicho en *pa*) + mundo posible
↓
Extensión
(V)
Cumplimiento de las condiciones de verdad de *pa*.

con
te
ni
do

Lo sostenido por *T* es verdadero, ya que corresponde con un hecho en el mundo, i.e. la altura de la torre Eiffel en $<w_{actual}, t_{1\text{proferimiento de pa}}>$. De hecho, la altura de la torre es de 324 metros, así 'lo dicho'[4] por *T* corresponde con cómo son las cosas en el mundo.

Veamos qué pasa con el enunciado expresado por *W*:

-Pa
La altura de la Torre Eiffel no es de 324 metros.

[4] 'Lo dicho' es aquello que es evaluado en la circunstancia de evaluación, esto es, la proposición.

↓
Establecimiento de las condiciones de verdad
Carácter $_{(f)}$: Contexto de uso <$w_{(a)}$,$t_{1(\text{proferimiento de-}\,pa)}$> →
contenido
Significado de *-pa* + contexto= intensión
Proposición *-pa*
Contenido $_{(f)}$
↓
Torre Eiffel (objeto relevante)
Altura de la Torre Eiffel (***Hecho en el mundo)***
Intensión (lo dicho en –*pa*) + mundo posible
con ↓
te **Extensión**
ni **(F)**
do Cumplimiento de las condiciones de verdad de *pa*.

En este caso, vemos que 'lo dicho' por *W* no corresponde con cómo son las cosas en el mundo, por ello es falso lo expresado. Tanto lo expresado por *T* (*pa*) como lo expresado por W (-*pa*), tienen el mismo *carácter*, pero tienen *contenido* distinto. Esto último se debe a que la intensión en 'lo dicho' por T, es distinta a la intensión en 'lo dicho por W, y es gracias a esto, que tenemos enunciados con un mismo tipo de significado i.e. *carácter*, pero con *contenidos* contradictorios, ya que tienen distinto valor de verdad. Por esto, podemos afirmar que sí hay un desacuerdo entre *T* y *W*, lo que faltaría es afirmar si tal desacuerdo es l*egítimo* o no.

Debemos rastrear las fuentes de la diferencia de opinión de *T* y *W*, que dan pie al desacuerdo. Para ubicar dichas fuentes, debemos tener información suficiente acerca de los *estados informativos* tanto de *T* como de *W*. Supongamos que tanto *T* como *W* adquieren la creencia de la altura de la Torre Eiffel de manera empírica. Ambos llevan consigo un flexómetro con el que miden la torre, y un cuadernillo en donde van sumando las distancias medidas. Ahora supongamos que *W* comete un error en una de las sumas, lo que causa que tenga como medida total de la altura de la torre Eiffel 314 metros. Claramente podemos ver que esto califica como un *error material*, ya que no depende del punto de vista particular de *W*, sino que se da con base un factor externo, esto es, un estado factual. Por tanto, podemos calificar el desacuerdo entre *T* y *W* acerca de (*pa*) como *legítimo*, ya que pudimos mostrar que se basa en una cuestión factual.

De este modo, podemos concluir que un desacuerdo se basa en enunciados objetivos, si en el argumento en disputa (ii) se comparte el *carácter*, pero se difiere en el *contenido*, teniendo como consecuencia valores de verdad distintos.

3. Desacuerdo, modelo de determinación de verdad y subjetividad.

Ahora supongamos que el sujeto Hopi está en desacuerdo con el profesor Willo acerca de la delicia del pulque de jitomate:

Willo: (pe) "el pulque de jitomate es delicioso."

Hopi: (*-pe*) "¡Claro que no! El pulque de jitomate no es delicioso."

Igual que en el caso anterior, primero deben establecerse las condiciones de verdad de la afirmación *pe.* Para ello deben cumplirse las condiciones (a y b) arriba expuestas. Ahora, la propuesta que hago es que los enunciados subjetivos difieren de los enunciados objetivos en la manera en que se establecen las condiciones de verdad del enunciado, y en el tipo de tópico (subject matter) al que estas condiciones apuntan.

La diferencia que propongo es que, para los enunciados que involucren juicios de gusto, debe de incluirse un tercer elemento para determinar el contexto de uso. Debido a que la verdad de estos enunciados depende casi totalmente del sujeto, se incluye a éste como elemento necesario para determinar el contexto de uso de los enunciados subjetivos, dando como resultado la triada <a,w,t>. Donde *a* representa el agente que profiere el enunciado en el contexto, *w* el mundo del contexto y *t* el tiempo del contexto; simbólicamente: a_c, w_c, t_c. De este modo, estos tres elementos intervienen en la fijación del *carácter* del enunciado subjetivo (*pe*). El *contenido* ubica los objetos relevantes para la evaluación de lo dicho, y para los enunciados subjetivos, el objeto relevante es el sujeto, específicamente, el gusto del sujeto. En estos enunciados la proposición tiene que ver con cómo son las cosas para el sujeto, específicamente con el gusto del sujeto.

A cada término singular del enunciado (*pe*) le corresponde un elemento de la proposición expresada en (*pe*). Ahora, 'el pulque de jitomate' se refiere a un objeto específico en un mundo, según nuestro ejemplo, el mundo en el que profesor Willo profiere (*pe*). Por su parte, 'delicioso' se refiere a un tipo de experiencia que el profesor Willo tiene al probar el pulque. Según lo arriba expuesto, el *carácter* del término 'delicioso' se determina a partir de un contexto que es determinado por la triada <a,w,t>. Esto tiene como consecuencia que el *carácter* de enunciados que contienen cuestiones de gusto dependa del sujeto que profiere tal enunciado, ya que el gusto es propio del agente. Al igual que Kaplan, pienso que en el *carácter* es donde reside la significancia cognitiva, a través de éste aprehendemos un pensamiento, esto es, individualizamos estados psicológicos[5].

El *contenido*, al ser el resultado de la función que es el *carácter*, será evaluado en una circunstancia de evaluación que también esté determinada por <a,w,t>. Así la proposición expresada en (*pe*) –

[5] Cfr. Kaplan David, 1989, <<Demonstratives>>, *In Almog, Perry, and Wettstein*, pp. 495 y 532.

contenido-, tomará como relevante la "vivencia específica" del profesor Willo, para ser evaluada, que en este caso es el gusto del profesor por el pulque de jitomate. Siguiendo el modelo arriba expuesto tenemos que:

Pe
El pulque de jitomate es delicioso
↓
Establecimiento de las condiciones de verdad
Carácter $_{(f)}$: Contexto de uso <a $_{(vivencia)}$,w$_{(actual)}$,t$_{1(proferencia\ de\ pe)}$> → contenido
Significado de *pe* + contexto= intensión
Proposición *pe*
Contenido $_{(f)}$
↓
Con te ni do
{Vivencia relevante+w+t}
Gusto del profesor + mundo + tiempo
Intensión + mundo posible+sujeto <a,w,t>
↓
Extensión
(V)
Cumplimiento de las condiciones de verdad de *pe*.

Vemos pues, cómo (*pe*) es subjetivamente verdadera, ya que es evaluada en una circunstancia que incluye al sujeto y a un aspecto relevante del mismo (vivencia).

Sin embargo, vemos que en lo expresado por Hopi, ocurre lo siguiente:

-Pe
El pulque de jitomate no es delicioso
↓
Establecimiento de las condiciones de verdad
Carácter $_{(f)}$: Contexto de uso <a $_{(vivencia)}$,w$_{(actual)}$,t$_{1(proferencia\ de\ -\ pe)}$> → contenido
Significado de *-pe* + contexto= intensión
Proposición *-pe*
Contenido $_{(f)}$
↓
Con te ni do
{Vivencia relevante+w+t}
Gusto de Hopi + mundo + tiempo
Intensión + mundo posible+sujeto <a,w,t>
↓
Extensión
(V)
Cumplimiento de las condiciones de verdad de *-pe*.

Por tanto, podemos concluir que no hay un *desacuerdo legítimo* entre el profesor Willo y Hopi, porque no se tiene una contradicción, ya que en ambos casos tanto *(pe*) como (*-pe*) son verdaderas. Tanto lo expresado en (*pe*) como lo expresado en (*-pe*), tienen distinto *carácter* y *contenido*. Además no podemos ubicar que las fuentes de diferencia de opinión se den con base en

algún estado factual. Por tanto, podemos ver que no hay realmente un desacuerdo, y, por ello decimos que el desacuerdo se pierde.

REFERENCIAS

Kaplan, D. 1989, 'Demonstratives', *In Almog, Perry, and Wettstein*, 481-563.
Kripke, S. 2005, *El Nombrar y la Necesidad*, México, Universidad Nacional Autónoma de México, IIF, 56-59.
Lasersohn, P. 2005, 'Context Dependence, Disagreement, and Predicates of Personal Taste', *Linguistics and Philosophy*, 28, 643-86.
McFarlane, J. 2007, 'Truth and Subjectivity', abril, *manuscrito inédito*.
Psillos, S. 2002, *Scientific realism. How science tracks truth*, Londres, ed. Routledge,. - Rodríguez Zaragoza, M. E. 2010, *Condiciones para la objetividad lógica. Caso particular: la noción de consecuencia lógica,* Tesis de Maestría, UNAM.
Shapiro, S. 2006, *Vagueness in Context*, Clarendon Press, Oxford University Press.
Skidelsky, L. 2007, 'La distinción doxástico-subdoxástico', *Crítica, Revista Hispanoamericana de Filosofía,* vol. 39, 115, 31-60.
Wright, C. 1986, `Inventing Logical Necessity', en J. Butterfield ed., *Language, Mind and Logic*, Cambridge University Press, 187-209.
- 2003, *Saving the differences. Essays on Themes from Truth and Objectivity*, Harvard University Press, Cambridge, Massachusetts, Inglaterra.
- 1992, *Truth and Objectivity,* Harvard University Press, Cambridge, Massachusetts, Inglaterra.

VOLUNTARY BELIEF ON A STRICTLY EPISTEMIC BASIS?

SERGI ROSELL
University of Sheffield
s.rosell@sheffield.ac.uk

Abstract: This paper discusses a recent proposal by Philip Nickel [2010], according to which we believe voluntarily—in cases of non-conclusive but adequate evidence—when the conditions of reason-responsiveness, alternative possibilities and self-expression are satisfied. I counter-argue, on the one hand, that Nickel's argument unjustifiedly conflates two different issues—the rationality of belief and the voluntariness of belief—of a quite distinct nature. Moreover, Nickel's interpretation of these cases is ad hoc, insofar as he assigns a role to epistemic character only for that kind of cases. Instead, I propose an alternative integrated explanation for those cases in the context of a general account of belief, which I sketch, which denies the possibility of voluntary belief of any sort.

Keywords: Doxastic Voluntarism, Rational Belief, Epistemic Character, Epistemic Reasons, Evidence.

1. The commonest defense of doxastic voluntarism insists on the possibility that when evidence for or against any proposition *p* is non-conclusive, then practical or pragmatic reasons can lead the subject to believe that *p* or not-*p*. In such cases, the subject would come to believe that *p* (or not-*p*) voluntarily. (James [1898], Ginet [2002]) However, Philip Nickel [2010] has recently argued for a new version of doxastic voluntarism on a strictly epistemic basis. According to him, in cases of adequate but non-conclusive evidence (evidential tie) it is reasonable or rationally permissible for the subject either to come to believe that *p* or to suspend judgment about *p*.

This is one of Nickel's target cases (313-4). Someone has lived for a long time near a place where there are train tracks, but he has never heard the sound of a train, until one morning in which he hears a train whistle and feels vibrations of the sort typically produced by a locomotive. In this situation, this person is in a position to voluntarily adopt the belief that there is a locomotive nearby, but it is also reasonable for him to suspend judgment in absence of more evidence. What would turn voluntary both forming the belief and suspending judgment is that the following conditions obtain:

(i) *reasons-responsiveness*: both coming to believe that *p* and suspending judgment would be responsive to reasons (in the sense of being those reasons the agent takes to be relevant to the belief that *p*, but also in the sense of being the real reasons that there are);

(ii) *subjective options*: there is more than one rationally permissible doxastic option genuinely open to the subject, so that they can opt for one state rather than the other (and no intermediate decision or specific mental act is required);

(iii) *self-expression*: the expression of one's epistemic character 'in her disposition to form one doxastic state rather than another' (318).

Crucially, conditions 2 and 3 are not to obtain in cases of conclusive evidence, since these are not cases of voluntary belief formation; whereas 1 obtains in all cases of rational belief, so is not a condition exclusive for voluntary belief formation.

Interestingly, Nickel's view is not pragmatic, insofar as there are no particularly pressing non-epistemic reasons for the subject to opt for one of the alternatives.[1] And he accepts the evidentialist's claim that *belief aims at truth*, which restricts the options available to the subject, so that 'conclusive evidence, when grasped by a doxastic subject, must induce belief.'

2. My reply to Nickel is twofold. On the one hand, his argument conflates with no justification two different issues: the *rationality of belief* (re which belief is reasonable or rationally permissible regarding available evidence) and the *voluntariness of belief* (re how we come to belief something). Once this distinction is made explicit, one can accept most of Nickel's proposal—it definitely seems right to claim that it is permissible for the subject both to belief that *p* or suspend judgment about *p* in these circumstances—but this does not involve accepting that this sort of belief formation is voluntary.

So, the main flaw in Nickel's argument is trying to explain the voluntariness issue in terms of what is rationally permissible. The debate about doxastic voluntarism is not primarily about justification, rationality or reasonability—what should one believe—but about the concept of belief and the phenomenology of (or rational reconstruction of the process to) coming to belief. This parallel should help: one issue is whether I should help you, and another whether I can. It seems clear that the first issue cannot settle the second.

This confusion is basically due to the fact that Nickel (as Clifford [1879] also did) takes evidence to mean objective evidence—or epistemic reasons as the reasons that there really are for believing that *p*. Even though he construes his condition of reason-responsiveness as relatively weak ['It is enough, for voluntary T, that she respond to the relevant reasons she takes there to be for T, and that she respond to at least some of the reasons there actually are for T' (321)], and this sounds

[1] It would be question begging against his view, says Nickel (322), to require for a belief to be voluntary acquired that it be based on practical reasons.

as a quite plausible condition for the issue of belief rationality; this construal misrepresents what is at stake in the debate about the possibility of forming a believe voluntarily. My contention here is that the right way to understand evidence regarding doxastic agency is as subjective evidence or, in other words, as evidence for the subject—i.e. that evidence accessible to the agent and how she grasps or interprets it.[2] Why this is so will hopefully become clearer after sketching my account of believing in the next section.

My second contention is that Nickel's account is ad hoc regarding the role assigned to the subject's epistemic character in belief formation—which he takes to play a role only in cases of evidential tie.[3] In order to show why this is a mistake, I will start by briefly sketching an account of belief—not intended as particularly controversial—which integrates cases of evidential tie in a broader framework. It seems to me that what is primarily (more basically) at stake in believing is not whether one believes or suspends judgment (important as this might be), but rather the degree of belief required by the evidence as grasped by the subject, and in this process one's epistemic character always has an (indirect) role to play.

3. Epistemic subjects assign a particular credit to every proposition that they consider. Considering involve many different things, ranging from a more or less automatic and tacit response (the extreme case being direct perception) to accurate and explicit reflection or process of inquiry. When a proposition *p* is presented to S's consideration S assigns it a certain probability to be the case. This probability is to be primarily understood in subjective terms, since it strictly depends on the evidence available to the subject. But whatever the probability that S assigns to *p*, S is always in a doxastic relationship to every proposition considered by her. Thus, the basic question is not whether one believes that *p* or not, but how likely one considers *p* to be. When assigned probability is positive, even if it is very slightly positive, we believe that *p*—although obviously the degree of

[2] Of course, from the subject's perspective, my evidence for (believing that) p is the evidence that there is for (believing that) p (transparency principle). But it also is possible for a subject to believe that p in virtue of certain epistemic reasons but at the same time doubt whether some of those reasons are good enough. Subjective and objective evidence always diverge to some extent (in non-omniscient agents), and criteria for rational belief may be more or less demanding on the adequacy of the former to the latter. In any case, the normative question must be distinguished from the explanatory or descriptive one.

[3] 'Cases in which there is rational latitude among doxastic options do provide a context for displaying certain epistemic character traits.' (320) He seems to follow Raz [1997] here. Additionally, Nickel claims that there are other, but also exceptional, contexts in which epistemic character traits can also play a role—those in which believing something would be devastating and the normal response, against reason, would be denial.

conviction will vary greatly between, for example, assigning that low probability or a very high positive one.[4]

No doubt, the degree of credit or probability that one assigns to both *p* and not-*p* may sometimes be very similar. Indeed, it can even be the same, what certainly would lead one to suspend judgment. It might also be slightly higher for not-*p* than for *p*, so that one will be slightly more inclined towards *p* than towards not-*p*. In this latter case, one would believe that *p*, though assigning to it very low credit. It seems then natural to think that in evidential tie cases it is not that one's epistemic character is exceptionally leading one to believe or suspend judgment. Epistemic character traits seem to be always involved in belief formation: according to one's ability to pay attention, openness to evidence, carefulness in reaching conclusions, reliability on their own inferences, etc., one assigns different credit to the considered proposition. No doubt, in cases of evidential tie, differences in epistemic character are likely to play a more dramatic role than in cases of conclusive evidence, since there epistemic personal traits can change or settle the polarity of belief (positive or negative); but this is not a reason to deny that they have a role to play in cases of conclusive evidence as well.

Epistemic traits, as Nickel (319) concedes, cannot function as reasons to believe. They rather mediate in the way the subject credits (weights, evaluates) evidence. It is not that in cases of non-conclusive-but-adequate evidence the subject contemplates the two options open (condition 2) and then opts for one of them. Instead, it is the particular nature of one's epistemic character what makes that one immediately interprets the evidence in favor of one or the other option, or of no-one. It might happen that the subject really contemplates both options as equally probable, but then, once arrived at this point, it seems difficult that, by making a step further, the subject finally might opt for one of them, without reconsidering (re-weighting or re-evaluating) the epistemic reasons available. If you really see the evidence as non-conclusive, you are simply not to be able to opt for any of the alternatives and will suspend judgment. In an important sense, we always have more than one epistemic option open when we begin to consider the evidence in question, but not once we have already evaluated and weighted this evidence. You can further come to accept one of the options, even if the probability assigned to *p* by the subject is close to, say, 0.5—and she then tends to suspend judgment—as the basis of inference or action policies, for non-epistemic reasons. Acceptance does not however alter the epistemic credit or probability that one assigns to *p*.

[4] Elsewhere, I have tried to put forward an analogy between believing and understanding. There are different ways of coming to understand, as there are different ways of acquiring a belief. Some are instantaneous and others require higher intellectual processing. It may be more or less difficult to understand something, but understanding occurs immediately once the evidence is properly grasped. See Rosell [2009].

4. *Conclusion*. Nickel's cases can perfectly be explained away. If they are not fundamentally different from cases of conclusive evidence, as I have argued, and in conclusive evidence cases we cannot claim that the subject comes to believe voluntarily, then we cannot claim that for evidential tie cases either. Belief always follows the same mechanism: assignment of credit in response to grasped epistemic reasons or evidence, a process in which epistemic character has always a role to play.

So, I have tried to show that, from the three previous conditions for voluntary belief, self-expression (iii) also obtains in cases of conclusive evidence, and then cases of evidential tie are not particularly voluntary; subjective options (ii) are either always open (before assigning weight to evidence) or never (after that); and reasons-responsiveness (i) clearly obtains in cases of conclusive evidence as well. Therefore, it is unjustified to conclude that cases of evidential tie are special in the sense of being cases in which the subject can voluntarily opt for believing or suspending judgment.

REFERENCES

Clifford, W.K. 1879, 'The Ethics of Belief,' in his *Lectures and Essays*, MacMillan and Co., edited by L. Stephen and F. Pollock (several re-editions).

Ginet, C. 2002, 'Deciding to Believe,' in *Knowledge, Truth and Duty*, edited by M. Steup, Oxford, Oxford University Press, 63-76.

James, W. 1987. 'The Will to Believe,' in his *The Will to Believe and Other Essays*, Longmans, Green and Co. (several re-editions).

Nickel, P. 2010, 'Voluntary Belief on a Reasonable Basis,' *Philosophy and Phenomenological Research* 81(2), 312-334.

Raz, J. 1999, 'When We are Ourselves: The Active and the Passive,' in his *Engaging Reason: On the Theory of Value and Action*, New York, Oxford University Press, 5-21.

Rosell, S. 2009, 'A New Rejection of Doxastic Voluntarism,' *Teorema* 28 (3), 97-112.

CARNAP'S PRINCIPLE OF TOLERANCE: PRAGMATISM OR NON-FACTUALISM?

J. SUROVELL
U. of Pittsburgh

ABSTRACT: Carnap's Principle of Tolerance holds that cognitive correctness does not enter into the choice of a language for science. For some, tolerance has significant appeal. However, tolerance's philosophical underpinnings remain far from clear. This essay aims to identify the philosophical basis for tolerance within Carnap's thought.

§1 more carefully characterizes the kind of metaphysical intrusion into language construction that tolerance is supposed to preclude. §2 proposes a pragmatist account of tolerance. Pragmatism holds that language is an instrument for the fulfillment of a chosen aim. For Carnap, the aim of scientific language is the description and derivation of observation reports. From the pragmatist perspective, then, the reason the choice of a language is not subject to cognitive correctness is that cognitive correctness does not affect the language's fulfillment of its function.

§3 contrasts my pragmatist interpretation with non-factualist interpretations of tolerance. According to non-factualist interpretations, tolerance holds that there is no coherent standard of cognitive correctness for language choice. I argue that non-factualist accounts do not address all of the metaphysical arguments that in §1 I argued fall within the ambit of tolerance. I conclude that pragmatism is the preferable understanding of tolerance.

Key words: Carnap, empiricism, analyticity, philosophy of language, semantics, principle of tolerance.

1 Introduction

Carnap's Principle of Tolerance holds that we are free to formalize our theories in whatever way we find most useful, as cognitive correctness does not enter into the choice of a language for science. For example, we may choose freely between nominalistic and platonistic theory formulations. Together with Carnap's conception of philosophy as the logical analysis of scientific theories, tolerance leads to Carnap's famous rejection of ontological debates as pseudo-disputes.

For some within the scientific philosophy tradition, tolerance has significant appeal. However, tolerance's philosophical underpinnings remain far from clear. In particular, Carnap's reasons for thinking language choice independent of cognitive considerations are not understood. This essay aims to identify the philosophical basis for tolerance within Carnap's thought.

I proceed as follows. §1 more carefully characterizes the kind of metaphysical intrusion into language construction that tolerance is supposed to preclude. §2 proposes what I call Carnap's 'prag-

matism' as an account of tolerance. Pragmatism holds that language is an instrument for the fulfillment of a chosen aim. For Carnap, the aim of scientific language is the description and derivation of observation reports. From the pragmatist point of view, then, the reason the choice of a language is not subject to cognitive correctness—the reason for tolerance—is that cognitive correctness does not affect the language's fulfillment of its function.

§3 contrasts my pragmatist interpretation with what I will call 'non-factualist' interpretations of tolerance. According to non-factualist interpretations, Carnap's tolerance holds that there is no coherent standard of cognitive correctness for language choice. I focus on the version of this view developed by Hylton. I show that Hylton's account does not address all of the metaphysical arguments that in §1 I argued fall within the ambit of tolerance. I conclude, on these grounds, that pragmatism is the preferable understanding of Carnap's tolerance.

2 The Nominalist

(Carnap, 1963) illustrates the kind of ontological approach to language choice that is at odds with tolerance through the example of two logicians, X_1 and X_2, who consider which language to use to formalize their theories. X_1, the relative platonist, opts for the language L_1, in which 'There are classes of classes' (or an equivalent symbolic formulation) is analytic. X_2, the relative nominalist, opts for L_2, in which 'There are no classes of classes' is analytic. The relative nominalist addresses the following argument to the relative platonist:

> In contrast to you, there is no possibility for me to choose between the two languages [L_1 and L_2]. On the basis of careful considerations I have arrived at the following two ontological results:
>
> (6) There are classes of objects.
>
> (7) There are no classes of classes of objects. (Carnap, 1963, 873)

In this passage, the relative nominalist argues as follows against adopting L_1:

> (N1) There are no classes of classes.
>
> (N2) L_1 is committed to the existence of classes of classes.
>
> Therefore,
>
> (N3) I cannot choose L_1 as my language for science.

Carnap considers the relative nominalist's argument to be fundamentally misguided. The nominalist is issuing metaphysically-driven prohibitions against certain language forms. Carnap's Principle of Tolerance denies that such prohibitions are ever legitimate. According to tolerance,

> *[i]n logic, there are no morals.* Everyone is at liberty to build up his own logic, i.e. his own form of language. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap, 1937, 52)

The freedom to use the language of our choosing asserted in this passage is freedom from what I will call 'cognitive correctness'; choices of a language for science are in some sense independent of "cognitive" or "theoretical" considerations like epistemic justification and representational adequacy. In place of cognitive correctness, we have only pragmatic considerations of expediency to guide our choice of a language. The nominalist's argument is at odds with tolerance in that it purports to identify a fact (the non-existence of numbers) that renders L_2 cognitively unfit for use in science.

Drawing on his (Carnap, 1956a) distinction between internal and external questions of existence, Carnap distinguishes two interpretations of (N1) and rejects the relative nominalist's argument on both interpretations. On the one hand, (N1) may be asked outside of a language for science, that is, independently of any logical or empirical rules that could help us to determine its truth value. Carnap calls the premise, when meant in this way, *external*. When (N1) is meant in this way, the relative nominalist's approach resembles what is sometimes called *first philosophy*—the relative nominalist attempts to establish the incorrectness of a scientific language by means of intuition or insight that is supposedly prior to science.

If, on the other hand, the relative nominalist intends (N1) as part of a language for science, then the premise is *internal*, and I will call the relative nominalist a *dogmatic empiricist*. The dogmatic empiricist is an empiricist in Carnap's (Carnap, 1936, 33) sense: all of her claims are either empirical or analytic within her language. She is dogmatic in that, as she contemplates changes to her language, she attaches some positive value to preserving her current postulates. For example, when faced with the option of adopting a new language L_1 that differs from his current language on the question of numbers, the relative nominalist considers L_1 incorrect *because* of this difference from his current language. The dogmatic empiricist's approach to language choice is captured by a *criterion of language correctness* discussed by Ricketts (Ricketts, 1994, 196): I consider a language correct to the extent that its theorems are all and only translations of my language's. The dogmatic empiricist need not regard the criterion of correctness as indefeasible, but it does exert some influence

on her decisions. The dogmatic empiricist might be motivated by the idea that "my conception of 'the way things really are' is represented by the language that formalizes my total theory" (Ricketts, 1994, 196), that "we have no higher access to truth than our evolving theory, however fallible" (Quine, 1992, 99).

Most commentators have focused exclusively on Carnap's rejection of external questions in their discussions of his critique of ontology. Though Carnap's focus in "Empiricism, Semantics, and Ontology" is generally the external questions, there are two textual reasons to think his critique extend to dogmatic empiricism as well. The first is his tendency to argue against the general approach on which linguistic prohibitions are decreed on the basis of ontological doctrine. Carnap voices this argument when he "[takes] the position that the introduction of the new ways of speaking does not need any theoretical justification because it does not imply any assertion of reality" (Carnap, 1956a, 214). In this typical passage, Carnap denies that the choice of a language is subject to *any* kind of putative cognitive justification for language choice; he does not deny only that external claims can justify the choice of a language. The second piece of textual support for my broader interpretation of Carnap's critique of ontology is to be found in his (Carnap, 1963, 873). There, Carnap explicitly objects to the nominalist's argument on the internal interpretation of (N1). And in addition to these textual motivations, there is the "spirit" of the principle of tolerance—surely Carnap would not accept the relative nominalist's argument even if (N1) were internal.

3 Pragmatism and Tolerance

(Carnap, 1956b) presents the following view of language:

> the choice of a certain language structure... is a practical decision like the choice of an instrument; it depends chiefly upon the purposes for which the instrument—here the language—is intended to be used and upon the properties of the instrument. (43)

The decision to use a language for science is the practical decision to use an instrument. Languages, like other instruments, have purposes for which they are to be used. The decision to use a particular language is therefore practically justified to the extent that the language achieves this purpose efficiently.

For Carnap, the purpose of scientific language is the description and accurate prediction of observable events. A language for science has three components, each of which makes a distinct contribution to achieving this aim. The observational sub-language is charged with the "description of observable events" (Carnap, 1956c, 41), including both recording observations and stating observable

predictions. The theoretical language axiomatizes general laws, including laws that posit unobservable entities. In the following passage, Carnap describes the purpose of a language's theoretical postulates:

> For an observer X to "accept" the postulates of T means here not simply to take T as an uninterpreted calculus, but to use T together with specified rules of correspondence C for guiding his expectations by deriving predictions about future observable events from observed events with the help of T and C. (Carnap, 1956c, 45)

Finally, the language contains logical and mathematical machinery, which "furnish instruments for deduction, that is, for the transformation of formulations of factual, contingent [i.e., empirical] knowledge" (Carnap, 1939).

Carnap's pragmatism is fundamental to his conception of scientific language. The familiar criteria for language choice that Carnap urges in a number of works, viz. expediency, fruitfulness, and simplicity, are the properties whose possession he believes best equips languages to carry out their descriptive and deductive tasks in empirical science. Empiricist languages—languages in which every sentence is either empirically significant or analytic—are to be preferred because the addition of descriptive, non-empirical vocabulary—non-observational vocabulary that makes no "difference for the prediction of an observable event" (Carnap, 1956c, 49)—would not improve the language *qua* deductive instrument for science.

More importantly for present purposes, pragmatism implies the principle of tolerance. Even if L_1's postulation of classes of classes were incorrect—in the sense of being false, or of the criterion of correctness discussed earlier—this would not render the relative platonist's decision to adopt L_1 less correct than the relative nominalist's decision to adopt L_2. According to Carnap's pragmatism, the choice of a language is to be evaluated for how well the language achieves the purposes for which it is to be used. On this view, there are purposes for which L_1 might be used such that the relative nominalist's argument will go through and other purposes such that the argument is invalid. In particular, the argument goes through if L_1 is to be used for some purpose for which the criterion of correctness is relevant.[1]

But given the aims for which the Carnapian uses scientific language, the relative nominalist's argument is invalid. If L_1's posit were incorrect, then at the cost of an incorrect theorem, the platonist's language would receive a "greater wealth in means of expression and deduction" (Carnap, 1963, 873) than the nominalist's. Indeed, from the Carnapian perspective, the incorrect postulate is not, in

[1]Thanks to Wayne Myrvold and Hartry Field for discussion of the points made in this paragraph.

and of itself, a cost; the correctness of a language's postulates is not even among the properties that, like simplicity, etc., are relevant to choosing a language. Having correct postulates does not, in and of itself, make a language a more effective instrument for describing or deriving observation reports. False postulates might be more likely to be disconfirmed. But from the perspective of Carnap's pragmatism, it is empirical inadequacy, and not falsity, that would count against a disconfirmed postulate. If adopting a false postulate would not imply the negation of a recorded observation report, but would simplify the theory, then the postulate would render the language a better deductive instrument for science, and Carnap's pragmatism would recommend adopting it.[2]

4 Non-Factualist Accounts of Tolerance

Up to this point, virtually all attempts to spell out tolerance's philosophical underpinnings advert to what I will call 'non-factualism'. According to non-factualist interpretations, there are no "language-transcendent facts" that would "[impose] constraints on the choice of a calculus, making some choices incorrect because those choices stipulate sentences to be true which are in fact false" (Ricketts, 2007, 210); there is no coherent standard of cognitive correctness for language choice.

Thus on Hylton's interpretation, tolerance "rests upon" an "ideal of rational adjudicability" (Hylton, 1982, 172), i.e., a distinction between "substantive" debates, where there is agreement "about what data would settle the issue" (Hylton, 1982, 173), and "verbal" debates, where there is no such agreement. The substantive debates are those that have been relativized to a language for science: "[f]or Carnap, language, and in particular logic, play a constitutive role in knowledge: only once a logic and a language are in place is knowledge possible. Only then is a rational dispute, or rational agreement, possible" (Hylton, 2007, 45).

Hylton attempts to derive tolerance from the ideal of rational adjudicability. The ideal of rational adjudicability holds that there can be a substantive debate only if a language is in place; for Hylton's Carnap,

> [i]t follows that the choice of language and logic, and questions which turn on that choice, cannot themselves be matters for such dispute: with no logic in place, there will be no agreed methods which might settle such a question. (Hylton, 2007, 45)

What is here said to follow—that there can be no substantive debate about the choice of a language—

[2]Thomas Ricketts has in conversation argued that Carnap accepts only language-relative concepts of truth, and that from this perspective, talk of false postulates is incoherent. The pragmatist need neither deny nor accept this view of the limitations of language-relative concepts of truth. Pragmatism secures tolerance even if language-relative concepts of falsity can apply to postulates. The pragmatist therefore takes on a lighter philosophical load than those who deny that language-relative falsity can apply to languages' postulates.

is the principle of tolerance.[3]

Hylton's inference from the ideal of rational adjudicability to tolerance is problematic. This move relies on a suppressed premise, *viz.* that language-relative standards of correctness are inapplicable to languages. Omitting this premise leaves open the possibility that correctness is language-dependent, as per the ideal of rational adjudicability, but that languages can be correct or incorrect relative to a language. The criterion of language correctness discussed above (p. 5) provides one way of spelling out this possibility: by means of this criterion, we can use our chosen language to substantively discuss the correctness of various languages. Hylton's non-factualism thus fails to account for tolerance's rejection of dogmatic empiricism; it has nothing to say about the relative nominalist's argument when (N1) is internal to a language for science.

Pragmatism is in a better position to deal with dogmatic empiricism. Instead of denying that there are coherent standards of cognitive correctness for languages, the Carnapian pragmatist claims that even if there were such a standard, it would be irrelevant to the decision to adopt a language for scientific purposes. This account excludes both the first philosopher's intuitive standard and the dogmatic empiricist's standard of correctness from language choice.

REFERENCES

Carnap, R. (1936, Jan.). Testability and meaning–continued. *Philosophy of Science 4*(1), 1–40.
Carnap, R. (1937). *The Logical Syntax of Language*. New York: Open Court. Trans. A. Smeaton.
Carnap, R. (1939). Foundations of logic and mathematics. *International Encyclopedia of Unified Science I*(3).
Carnap, R. (1956a). Empiricism, semantics and ontology. In *Meaning and Necessity: A Study in Semantics and Modal Logic* (Second ed.)., pp. 205–221. Chicago: University of Chicago Press.
Carnap, R. (1956b). *Meaning and Necessity: A Study in Semantics and Modal Logic* (Second ed.). Chicago: University of Chicago Press.
Carnap, R. (1956c). The Methodological character of theoretical concepts. *Minnesota Studies in the Philosophy of Science I, The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, 38–76. H. Feigl and M. Scriven, editors.
Carnap, R. (1963). Replies and systematic expositions. In P. A. Schilpp (Ed.), *The Philosophy of Rudolf Carnap*, pp. 859–999. LaSalle, IL: Open Court.
Friedman, M. (2001). *Dynamics of Reason*. Stanford, CA: CSLI Publications.
Hylton, P. (1982, Aug.). Analyticity and the indeterminacy of translation. *Synthèse 52*(2), 167–184.
Hylton, P. (2007). *Quine*. New York: Routledge.
Quine, W. (1992). *Pursuit of Truth*. Cambridge, MA: Harvard University Press.
Ricketts, T. (1994). Carnap's principle of tolerance, empiricism, and conventionalism. In P. Clark and B. Hale (Eds.), *Reading Putnam*, pp. 176–200. Oxford, UK: Blackwell.
Ricketts, T. (2007). Tolerance and logicism: Logical syntax and the philosophy of mathematics. In

[3](Friedman, 2001, 31) and (Ricketts, 2007, 208) view tolerance as arising from the language-relativity of correctness in essentially the same way.

M. Friedman and R. Creath (Eds.), *The Cambridge Companion to Carnap*, pp. 200–225. New York: Cambridge University Press.

IS NATURALISM NORMATIVE? SOME PROBLEMS WITH NATURALIZING EPISTEMIC NORMATIVITY

BARBARA TRYBULEC
UMCS

ABSTRACT: In the paper I pose a question whether naturalized epistemology can be conducted as a normative enterprise. It is noticeable that naturalists cannot deliver substantial, traditionally understood account of normativity. This is the reason why naturalism is often conceived as a merely descriptive enterprise. Nevertheless, naturalists do not reject normativity yet they understand it in a very specific sense. In the paper I show the main problem with naturalising epistemic norms which could be expressed in the question: What is the difference between norms and description of facts? My thesis is that naturalistic epistemic norms are actually descriptions of empirical phenomena which are expressed in a normative form for the sake of everyday life and scientific practice.
KEY WORDS: naturalized epistemology, epistemic norms, epistemic justification, naturalistic fallacy.

Introduction

One of the most significant features of recent epistemology is the turn toward naturalism. Although there are many different explications of this turn one can find a few theses which clearly differs naturalists from "traditional" epistemologists. This paper attempts to point these of them which are concerned with the problem of normativity. I undertake the following issue: can naturalism be developed as a normative enterprise? In other words, what conditions must it satisfy to achieve a status of epistemology? My strategy consists in showing the standard solution to this problem which I call "the standard condition" and in analysing whether it can be satisfied by naturalism. According to "the standard condition", in order to give a substantial account of normativity naturalism must present a theory of epistemic norms which are derived from descriptive statements about facts but which are not replaceable with them. This claim is widely shared among naturalists. In spite of this, most of the positions within naturalized epistemology have problems with satisfying "the standard condition" for it is often understood too strongly and "traditionally" for naturalism. Later on I show that although naturalism cannot present the conception of genuine epistemic norms which satisfy "the standard condition" it is still normative. The suggestion is that

within naturalism epistemic norms are descriptive statements about empirical phenomena which are expressed in a normative form for the sake of everyday life and scientific practice.

Normativity within naturalism is understood in a different sense from that in traditional, aprioristic epistemology. The difference is so significant that using the term "normative" during the discussion often leads to confusion. The reason it that naturalists change meanings if the main evaluative epistemic terms such as "epistemic norm" and "doxastic justification, more precisely, they change conditions under which beliefs are regarded to be justified. This situation leads to numerous misunderstandings between disputants unless they are aware of the differences and the reasons for their occurrence. Therefore, it seems necessary to specify what "normative", "epistemic norm" and "justified belief" mean in naturalism in order to answer the question whether naturalism is normative according to its own standards and not traditional. Thus, I firstly recall the traditional, Cartesian meaning of "normativity". Secondly I show how this meaning has changed within naturalism and to make this difference clear I will refer to Quine's and Goldman's epistemology. Thirdly, on this basis I show the main problems with the process of naturalizing normativity and the way they can be overcame.

1. Epistemology as a normative enterprise

Why epistemology is claimed to be a normative enterprise? Traditionally knowledge is defined as a true, justified belief. Normativity is introduced to this definition by the term "justified", for "true" and "belief" are not strict epistemic terms[1]. "Justification" is a normative concept and it makes a discipline that builds its theory normative. Namely, from epistemic point of view, if a belief is justified it is valuable. The aim of the traditional epistemology rooted in Cartesian philosophy is to specify the criteria of justified belief that allow us to evaluate which beliefs we should accept and which reject. Epistemologists' task therefore is to explain when a belief is justified, what qualities it should have and what a subject should do to posses a set of true, justified beliefs called knowledge[2]. In spite of the significant differences between traditional and naturalistic epistemology, the normative task of this discipline reminds the same: to show how a subject should act to achieve knowledge, that is, the set of valuable beliefs.

[1] Kim 1988. "Belief" is a psychological term and "true" is a semantical – metaphysical one.
[2] Ibid.

2. Meaning of normativity within traditional epistemology

The first step in every discussion on normativity should consist in specifying one's attitude to the naturalistic fallacy, namely to the problem of deriving normative statements from descriptive ones. This attitude is the basic assumption of every theory of epistemic norms, for it determines the position on the sources of these norms and, consequently, on the criteria of a justified belief.

The standard position among traditional epistemologists is that epistemic norms are derived from and justified by some basic, prior source. This source is independent of every belief and it does not need any further justification, for it is evidently and perfectly reliable. Epistemic norms derived from such a source are universal, and they form the foundation of whole knowledge about the world. This is the Cartesian structure of knowledge based on infallible fundament[3].

In recent epistemology the Cartesian project is continued by internalists with respect to justification. The representatives of this position are Roderick Chisholm and Laurence BonJour[4]. Consequently, the subject must be aware of reasons which justify his belief; if this condition is not satisfied a belief is not justified[5].

Such a position on justification is based on the specific attitude to the naturalistic fallacy. Internalists argue that the derivation of normative statements from descriptive ones is a logical and definitional error. Norms should possess universal, normative force and therefore they should be justified with reference to an aprioristic source and not to knowledge about facts that humans already possess. Thus, the way in which a subject should form his beliefs has little in common with the way in which he actually forms them. It was unthinkable for epistemologists continuing Descartes' heritage to justify a belief by reference to cognitive processes which caused it, in other words to treat a cause as a reason, something which is so commonsensical and obvious for naturalists. This is a clear example of how different position can emerge from disagreement on one assumption, in this case on the assumption about naturalistic fallacy.

3. Meaning of normativity within naturalised epistemology

The critique of traditional epistemology conducted from naturalistic point of view is, in the most of cases, based on the one, central objection: epistemic norms justified in aprioristic manner cannot form necessary and infallible foundation of knowledge about the world[6]. According to naturalists, it is very difficult to specify criteria of evidency or indubitability – features that epistemic norms

[3] Descartes 1970
[4] BonJour 1985
[5] Many naturalists cannot agree with this statement. See: Kornblith 2002
[6] Knowles 2003

derived from infallible source should possess – which most philosophers would accept. How to justify these criteria in non-regressive way? Are there any epistemic norms that could meet them? How indubitable, aprioristic norms exist, and how a subject can grasp them[7]. In other words, how it is possible not to make naturalistic fallacy? For there are no satisfying answers to these questions, naturalists do not see other possibility then to derive epistemic norms from empirical information already possessed by humans. Naturalistic fallacy is for them no fallacy at all or this is a fallacy that cannot be avoided. Naturalists agree that it is logically impossible to derive normative propositions from descriptive ones, however this logical principle is useful neither in scientific practice nor in commonsensical thinking. In both areas of human activity epistemic norms which form criteria of a justified belief are rooted in descriptive knowledge about facts.

Traditional philosophers do not see any possibility for naturalism to be an epistemology, for the theory which commits one of the worst philosophical fallacies cannot be called normative. Rejecting apriority however, naturalist do not reject the idea of normativity, namely, the idea that there are reasons why one belief is better justified in some circumstances then the other one. Moreover, many naturalists claim that they solve normative problems in a more effective manner. This optimism is justified by the fact that they evaluate subject's cognitive abilities with relation to the particular environment and not with relation to the logical principles that are correct in every possible world[8]. This gives them the advantage of giving very specific advices how to form justified beliefs in particular circumstances. Roughly speaking, the task of naturalized epistemology is to evaluate which among humans' cognitive processes described by sciences are rational and why.

4. Quine's and Goldman's understanding of epistemic normativity

The change of meaning of "normativity" and "epistemic norm" can be seen clearly in Quine's epistemology. This author presents a radical form of naturalism which seems to reject the whole project of normative epistemology focused on justification. This rejection is motivated by the claim that epistemology is no longer The First Philosophy with its own autonomous method. Epistemologists must refer to science in order to evaluate a belief, however, science examines only causal relations between cognitive processes and beliefs. Thus, the normative function of epistemology is reduced to minimum[9]. It does not criticize scientific achievements and does not built foundation of science but it assumes the validity and warrant of scientific project and it uses scientific achievements in order to formulate the criteria of justified beliefs.

7 Ibid.
8 Kornblith 1994
9 Quine 1969

Examining Quine's epistemology, one can clearly notice how he changes the traditional meaning of "an epistemic norm". From an infallible, aprioristic foundation it transforms to empirical information about correct processes of reasoning. A sentence which seems to express a norm has only an advisable structure but actually it describes relations between facts, namely, it shows which cognitive processes humans use to gain knowledge, which processes are reliable. "The standard condition" stated in the beginning of this paper is therefore not satisfied by Quine's naturalism.

Alvin Goldman is another inspiring epistemologist who tries to solve the problem of naturalistic normativity. According to Goldman's reliabilizm, a belief is justified if it is a product of reliable process, namely a process that leads in the majority of cases to the truth[10]. Justification of a belief dependents therefore on appropriate causal relation between a subject's belief and a process which caused it. Contrary to internalists, representatives of this theory claim that a subject does not need to be aware of the process which justifies his belief, this process could be wholly external to the subjective evaluation of a justificational status of this belief[11]. As a result, a subject does not longer play leading role in the process of justification, he does not need to consciously obey certain epistemic norms for he uses psychological, cognitive processes which could proceed beyond his consciousness. Hence, it seems that epistemology is redundant, for humans do not need to make special effort and consciously obey some epistemic norms in order to form rational beliefs, they just do it by nature for this is the way in which their cognitive system works.[12]

To sum, naturalizing justification could be understood as the process of changing an accent from the logical analysis of relations between propositions to the analysis concerned with cognitive processes which form beliefs. If these processes are correct they make a belief justified. Hence, traditional reasons have changed within naturalism into causes and aprioristic epistemic norms which built foundation of knowledge have changed into statements describing how actually cognitive processes are performed by humans.

5. Some problems with "the standard condition"

It is claimed that if naturalism wants to be called epistemology it cannot loose normative character. Therefore, it has to propose a theory of justification based on epistemic norms which are derived from factual statements but are not replaceable with them. In other words, epistemic norms should be different in a substantial way from descriptions of facts. If this condition is not satisfied, there

[10] Goldman 1986
[11] Goldman 1979
[12] Knowles 2003

could appear a danger of losing by naturalism the possibility of analyzing the problem of justification, for the term "justification" is normative. Naturalism however does not see any problem with deriving normative conclusions from the empirical knowledge. Is there really no problem with that? How naturalists can differentiate between norms derived from factual statements and these statements? What is the function of epistemic norms which descriptions of phenomena cannot fulfill? The crucial problem for naturalism is therefore not whether one can derive normative from descriptive but whether there is something like real normativity at all and what is the source of it.

If epistemic norms are reducible to empirical facts how to make sense of norms which are in some substantial way different from mere description of relations between facts? Norms seems to be redundant, for subject in order to possess knowledge has to refer to these descriptions instead of norms[13]. Thus, epistemology looses its subject, for if our cognitive practices are fundamentally correct there is no need to form norms which would tell us how we should cognate. Consequently, naturalism cannot present a substantive account of normativity and it cannot be an epistemology. Yet is it really so?

6. Naturalism without "genuine" normativity

The crucial question is why naturalism must satisfy traditional standards? If naturalism cannot deliver genuine epistemic norms (norms in the traditional sense) does it mean that it cannot present any theory of epistemic norms? How to make sense of naturalized epistemology that do not refer to the traditionally understood genuine epistemic norms?

Traditional expectation of real, substantive normativity is too high and completely unrealistic for naturalistic presuppositions[14]. The suggestion is therefore to understand naturalistic normativity in a very reduced and minimal manner but which is, however, enough for naturalised epistemology. How it could be done?

[13] See Knowles 2003

[14] J. Kim maintains that epistemic qualities which are among others norms are not reducible to facts but they supervene on descriptive properties of phenomena. Knowles would probably ask why we need such kind of supervening norms, for if they rest on facts so why do not simply refer to facts? See Kim 1988.

7. Pragmatic account of naturalised normativity

Naturalistic account of normativity cannot be accepted by traditionally oriented philosophers. However, naturalists, do not see in this project nothing controversial for it is everything they can afford being faithful to their naturalistic presuppositions. The account of normativity which is suggested here could be called pragmatic, for epistemic norms within naturalism are actually descriptive, empirical statements which are treated as norms for pragmatic reasons. Thus, there is no real derivation between norms and factual knowledge, namely norms are not derived as something substantially different. An epistemic norm is the reformulation of a factual statement needed in advising what the most effective way of reaching some cognitive aim is. Naturalised normativity is much weaker than traditional in the sense that it does not present infallible and unchangeable standards. Naturalism "reduces" – traditionally saying – epistemic norms to unstable descriptions of empirical phenomena. From traditionalists' point of view it means that naturalism is descriptive and it cannot conduct epistemological research. One cannot however demand, from naturalism to satisfy traditional standards of a substantial account of normativity. Instead, it should satisfy its own standards, namely to point out, on the bases of scientific research, what conditions valuable, rational beliefs should satisfy.

Summarizing, naturalism can be conducted as a normative enterprise even if it does not satisfy traditionally understood "standard condition". Epistemic norms are descriptive propositions which are treaded as norms and express in a normative way for the sake of everyday and scientific practice. The naturalized epistemology is normative as far as it could be according to its own presuppositions. It forms conditions of justified belief on the bases of scientific descriptions of humans' cognitive processes and abilities. It is essential to remember not to confuse traditional and naturalistic meanings of epistemic concepts and standards since these two epistemological positions are formed from very different philosophical standpoints. Forgetting about this leads to misunderstandings and confusion which make impossible any fruitful discussion between the two camps.

REFERENCES

BonJour, L., 1985, The structure of empirical knowledge, Cambridge, Harvard University Press.
BonJour, L. 1994, 'Against naturalised epistemology', Midwest Studies in Philosophy, XIX, 283-300.
Chisholm, R., 1977, Theory of knowledge, 2nd edn., Prentice-Hall.
Descartes, R. 2010, Discourse on The Method, http://records.viu.ca/~johnstoi/descartes/descartes1.htm

Goldman, A. 1979, 'What is justified belief', in. Kornblith H. 1994 (ed.), Naturalizing epistemology, 2nd edition, MIT Press, Cambridge, Mass, 105-130.
Goldman, A., 1986, Epistemology and Cognition, Cambridge, Harvard University Press.
Kim, J., 1988, 'What is "naturalised epistemology"?' in Kornblith H.1994, (ed.), Naturalizing epistemology, 2nd edition, MIT Press, Cambridge, Mass, 33-55.
Kitcher, P., 1992, 'Naturalists return', The Philosophical Review, Vol. 101, No. 1, 53-114.
Knowles, J., 2003, Norms, Naturalism and Epistemology, Palgrave Macmillan.
Kornblith, H., 1993, 'Epistemic normativity', Synthese 94, 357-376.
Kornblith H., 1994, 'Naturalism: Both Metaphisical and Epistemological', Midwest Studies in Philosophy, XIX, s. 39-52.
Kornblith, H., 1994, 'Introduction: What is naturalistic epistemology?' in Kornblith H. 1994, (ed.), Naturalizing epistemology, 2nd edition, MIT Press, Cambridge, Mass, 1-14.
Kornblith H., 2002, Knowledge and its Place in Nature, Oxford, Clarendon Press.
Laudan, L., 1984, Science and Values, Berkeley, University of California Press.
Quine, W., 1969, 'Epistemology Naturalised' in Kornblith H. 1994, (ed.) Naturalising epistemology, 2nd edition, MIT Press, Cambridge, Mass, 15-31.
Quine, W., 1986, 'Reply to Morton White', in Hahn L., Schilpp P. (eds.), The philosophy of W. V. Quine, La Salle: Open Court, s.664-665.
Salmon, W. 1998, Causality and Explanation, New York, Oxford University Press.

UNBELIEVABLY GOOD NORMS

VÍCTOR M. VERDEJO
Universidade de Santiago de Compostela

Abstract: Recent discussion on the normativity of content has taken the normativity of (true) belief to be one, if not the most salient, of its manifestations. In spite of an overwhelming consensus, I argue, the normativity of belief is largely orthogonal to the assessment of the normativity of content. Indeed, our best theories of the normativity of mental content, that is, Inferential Role Semantics, are naturally understood as showing no commitment towards norms for (true) believing. This point can be used to offer powerful replies to anti-normativist criticisms. Here I discuss one central example.
Keywords: normativity of belief, normativity of content, IRS, normative requirements, regress of beliefs

1. Believably bad norms

Philosophical discussion has lately put normativity on the spot. In something like an anti-normativist crescendo many authors have tried to undermine what was, until recently, the dominating Kripkean dogma (Kripke 1982): meaning and content are inherently normative. Following standard formulations, the claim that meaning and content are inherently normative is the claim that the meaning of an expression, or the content of the thought it conveys, is or constitutes a genuine obligation. In other words, if the normativity view is correct, specifications of meanings or contents turn on specifications of how one ought to use a certain expression or how one ought to employ a certain concept or thought.

Many authors have defended that neither meaning nor content are inherently normative (Bykvist and Hattiangadi 2007; Hattiangadi 2006, 2009; Glüer and Wikforss 2009; Miller 2008). Strikingly, when assessing the normativity of mental content, both critics and supporters of normativity have showed a firm preference towards approaches that do not focus on content itself but, quite differently, on belief. As a quick look at the literature can show, the main (and often the only) focus of attention has been a principle along the lines of True Belief (TB):

(TB) For any *p* and *S*: *S* ought to believe that *p*, if and only if *p* is true.

Analyses of the variants and validity of this general principle abound in the literature (Boghossian 2003, 2005; Bykvist and Hattiangadi 2007; Miller 2008; Horwich 2006; Glüer and Wikforss 2009). In the light of this, one may very well wonder what would explain the fact that "attempts to show

that content is normative typically go via the idea that *belief* is essentially normative" (Glüer and Wikforss 2009, 33-4, emphasis theirs). In other words, *why* do philosophers spend their efforts and time offering analyses of (TB) as relevant analyses for the assessment of the normativity of content? Whereas I lack a definitive answer to this question, here is a pair of hints that probably point in the correct direction. In the first place, (TB) is an intuitively plausible principle: it really seems sound to assume that true belief is correct belief and, to that extent, normatively demanding (Wedgwood 2002, Gibbard 2005). In the second place, it seems as though we lack, really, an alternative to (TB) to stating, generally, what it would be for mental content to be normative. The leading thought is therefore that there is no clear way for laying out the normativity of content directly, and that we might have to opt for some indirect route (Boghossian 2003, 2005).

Some straws in the wind suggest, however, that it is no good idea to rely on (TB), or a principle in its vicinity, for a defence of the normativity of content. For instance, (TB) has at least two intriguing features. First, it states a normative relation to something that is blatantly *not* normative, namely, truth. As Gibbard once put it, "to say that *p* is true is a non-normative claim if any claim is non-normative" (Gibbard 2005, p. 38). Thus, it should strike one as surprising that intrinsic normativity could ever be stated or derived from a relation towards something that is clearly not normative, let alone intrinsically normative. On the other hand, (TB) is very obviously *not* a thesis about (propositional) content, but about a (content-involving) propositional attitude. (TB) seems to pose therefore an uncomfortable double task for the normativity defender: first, argue for the correctness of (TB); second, show that the truth of (TB) turns on the truth of a normativity thesis about content. Not an easy working day, indeed.

Surely because of these intriguing features, (TB) has been severely put into question in recent developments (Bykvist and Hattiangadi 2007; Miller 2008; Glüer and Wikforss 2009). I am inclined to think that these criticisms are largely correct as far as they go. Nonetheless, I dare to say too, they are also a wonderful instance of *ignoratio elenchi* when considered as criticisms of the normativity of content. The reason is that (TB) is, to repeat, patently not a thesis about the normativity of content and certainly not the less controversial instance of an intentionality-involving normative principle.

There is an alternative to (TB) that is worth considering and whose proper formulation has, at least in principle, nothing whatever to do with a normativity issue about true belief. The central idea of this alternative, which I now advance for the record, shall be that, whereas mental content

essentially is or consists of a series of substantial norms –even substantial norms involving beliefs–, it very much never involves norms *to believe* so and so.

2. Unbelievably good norms

Inferential Role Semantics (IRS henceforth) is by far the most natural place to look at when pursuing the normativity of mental content. Peacocke (1992) or Wedgwood (2007) are central examples of this kind of theory –a kind that results in what I call elsewhere the Rationality Framework (Verdejo 2009a, 2009b, forthcoming). I do not mean to suggest that this kind of theory has not been considered at all in the normativity debate. However, when it has been (e.g. Horwich 2005; Glüer and Wikforss 2009) discussion has also been disappointingly infected by reflection on (TB). Here I wish to show that IRS provides support to a form of normativity that is not of the (TB)-sort. More precisely, I point out how IRS yields an alternative principle for the assessment of the normativity of content.

The most elementary way to see the plausibility of this alternative principle is to note two fundamental aspects of IRS. The first aspect is that, when a full and correct IRS statement of a conceptual or propositional content *p* is available, this statement individuates *p*. The second aspect is that, when such a statement is available, it comprises a *possession condition* that specifies what is required for any thinker to actually be a *p*-thinker. What IRS provides therefore is what the normativity-of-content thesis calls for: a kind of mental content that is normatively articulated. Following this, we can have a feel for the correctness of Normativity of Content (NC):

(NC) For any *p* and *S*: if *S* thinks that *p*, then *S* ought to meet condition *C*(*p*).

Where *C*(*p*) both individuates content *p* and presents an obligation for *p*-thinking. To be sure, the anti-normativist will try to provide a non-normative version of (NC). In this paper, however, I am not concerned with showing the correctness of (NC). My aim is more modest: to show that if (NC) is taken at face value, then the normativity of content is very clearly something over and above the (TB)-normativity of belief. The fundamental point is that the (NC)-analysis is directly an analysis of mental content, and not of some not directly contentful element, such as a belief-attitude. Let me illustrate this with an example.

For the sake of the argument, let us assume that the correct IRS analysis for the concept *red* –*C*(*red*)– is the following: in order for *S* to possess *red*, when the circumstances are appropriate, *S* must find primitively compelling transitions of the form: $red(x)/\rho(x)$, $\rho(x)/red(x)$, where $red(x)$

stands for a particular (propositional) application of the concept to an object x and $\rho(x)$ stands for the (propositional) attribution of a particular phenomenological property to x.[1]

If $C(red)$, as here stated, were correct, it would, on the one hand, individuate the concept *red* via an inferential role and, on the other, introduce a general obligation for *red*-thinkers: if one is to think *red* (as opposed to *green*, *carburettor* or *risk premium*) one must find instances of the inferential role compelling. Now, this is only one example of zillions that would satisfy two features that I discuss in turn.

First, on this account, to think *red* is not necessarily a matter of having beliefs containing *red*. This is so even if, more often than not, beliefs are our best guide for coming up with satisfactory inferential roles. The point is that, very clearly $C(red)$ can be satisfied by, e.g., desire-transitions, fear-transitions, expectation-transitions, etc., in addition to belief-transitions. Indeed, I dare to say, $C(red)$ constrains even mere entertaining- and imagining-transitions. Thus, all of the following are clear instances of $C(red)$. For any S, if S thinks *red*, then S ought to:

desire $red(x)$, if S desires $\rho(x)$ and desire $\rho(x)$ if S desires $red(x)$;

expect $red(x)$, if S expects $\rho(x)$ and expect $\rho(x)$ if S expects $red(x)$;

imagine $red(x)$, if S imagines $\rho(x)$ and imagine $\rho(x)$ if S imagines $red(x)$.

To illustrate, plausibly, one ought to (or must) be compelled to accept that if one desires the traffic lights to be red, then, according to the propounded $C(red)$, one desires the traffic lights to have a certain phenomenological property. Similarly, one ought to (or must) be compelled to accept that if one imagines the grass to be red, then one imagines it to have a certain phenomenological property. And so on.

This shows that, *pace* Boghossian (2003, 2005), the normativity of belief may not have such a fundamental role regarding the statement of normativity. More precisely, what this shows is that IRS analyses are analyses of content across a variety of attitude-transitions, and not specifically for belief-transitions.

Secondly, and this is definitive for our purposes, *even* in the case in which IRS-analyses provide norms in the form of belief-transitions, they, very plausibly indeed, never involve norms for believing a certain proposition. What IRS analyses yield are not obligations to have a belief, or any other obligation to have a given attitude towards a propositional content. On the contrary, IRS analyses are naturally conceived as providing obligations to respect a certain inferential role. This is

[1] On this reading C(*red*) involves propositional transitions. This is a much simplified version of Peacocke's account of the concept (see Peacocke 1992). The example shows that one does not need to appeal to logical or some basic rationality-grounding concepts for the argument to go through.

so even in the case in which the role in question concerns transitions among beliefs. But an obligation to respect a belief-involving inferential role does not amount to an obligation to believe. The difference between obligations to believe that *p* and obligations to respect a belief-involving role can be put in terms of John Broome's (1999) difference between detachable and non-detachable normative relations. Among the latter we find the relation of *normative requirement*. According to Broome, to say that *p* normatively requires *q* is to say that one ought to (see to it that) if *p*, *q*:

$$(1)\ \mathrm{O}\,(p \rightarrow q)$$

Broome distinguishes normative requirements from the 'oughts' relation. When *p* oughts *q*, one ought to *q*, if *p*.

$$(2)\ p \rightarrow \mathrm{O}q$$

Unlike (2), (1) does not mean that if *p*, then one ought to (see to it that) *q*. That is, "a normative requirement does not permit a normative conclusion to be detached by modus ponens" because "the normativity is attached *to the relation* between the propositions *p* and *q*" (Broome 1999, 402, my emphasis). My point is that IRS are naturally interpreted as providing a sort of normative requirement among (belief-) transitions and not some other sort of detachable normative relation. At most then, what IRS provides are combinations (of beliefs) one ought to see to, but very much not single obligations to believe a certain proposition.

3. Blocking the regress of beliefs

It seems to me that an IRS approach to the normativity of content provides us with compelling lines of response against many anti-normativist objections that capitalize on the ambiguity of normative principles (Hattiangadi 2006; Glüer and Wikforss 2009) or their (psychological or logical) non-satisfiability (Boghossian 2005; Bykvist and Hattiangadi 2007; Hattiangadi 2009; Glüer and Wikforss 2009). In this paper, I wish to deepen the analysis of a more recent anti-normativist argument, namely, the one that appeals to regresses of rule-following (Glüer and Wikforss, 2009, 55-9). The argument is supposed to apply to normative views across the board. Here, I will concentrate on the simplest, belief-version. This is the objection: for (TB), or principles of its kin, to be genuinely normative subjects must carry out or have available a practical reasoning that involves believing that a belief that *p* is in accordance with the norm. More precisely, if one ought to believe that *p*, it must be that one is (possibly) motivated to follow the rule *R* that specifies the obligation to believe that *p*. But any such motivation requires a practical reasoning of the form:

(P_1) I want to believe what is in accordance with *R*.

(P_2) To believe that *p* is in accordance with *R*.

(C) I want to believe that *p*.

However, so Glüer and Wikforss argue, this leads to a regress of motivating beliefs. The reason is that one such reasoning would involve another belief, namely, the belief that (P_2) –viz. the belief that to belief that *p* is in accordance with *R*–, and one must explain, for a full account, how one is motivated to have that belief. The answer is: with yet another practical reasoning which would include yet another "doxastic slot" (Glüer and Wikforss 2009, 56). A vicious regress of normativity-motivating beliefs ensues.

Now, one may dispute that genuine normativity requires a practical reasoning of the form presented by Glüer and Wikforss (Jarvis 2012). However, in my view, the account I have provided unfolds a far-reaching response. According to the IRS approach I have been presenting, the regress of normativity-motivating beliefs does not as much as arise in the first place. Consider the following principle based on one of the target transitions of our *red* example above:

(RED) If *S* thinks *red*, then *S* ought to believe *red*(*x*) if *S* believes $\rho(x)$.

I am assuming, for the sake of the non-normativist argument, that the target transitions in (RED) are actually belief-transitions. This might be disputed but it seems to me that one must acknowledge that in many central cases the target transitions would actually have to be belief-transitions. Now, the point is that from (RED), no obligation to believe a given proposition follows. To repeat, even if (NC)-normativity plausibly involves plenty of obligations or prescriptions regarding beliefs, it very likely never involves a general obligation to believe or to have any other attitude regarding a proposition *p*. To illustrate, (RED) may have one of the two following readings:

(RED*) If *S* thinks *red*, then *S* ought (to believe *red*(*x*)) if *S* believes $\rho(x)$.

(RED**) If *S* thinks *red*, then *S* ought (to believe *red*(*x*) if *S* believes $\rho(x)$).

As the reader may expect, the difference between (RED*) and (RED**) is the difference between a detachable and a non-detachable normative relation. The key point is that since what IRS provides are inferential roles –i.e. combinations of attitudinal, and perhaps perceptual and other psychological states– we should prefer (RED**) over (RED*) as a relevant analysis of (RED). But if this is right, note that Glüer and Wikforss's regress of beliefs does not even get started. When the 'ought' takes wide scope, as in (RED**), the normativity in question involves no obligation to believe or to have pro-attitudes towards a given content, but quite differently, it concerns obligations to fulfil or respect a certain psychological role. If this is correct, Glüer and Wikforss's

regress is blocked from the start even if we accept, as we might not, that genuine norms or rules involve a form of practical, motivating reasoning.

Conclusion

If the view here presented is sound, it would seem that our best candidates for the defence of the normativity of mental content (viz. IRS theories) are not, after all, in the least committed to a thesis about the normativity of (true) belief. The moral should not, on reflection, be surprising. Usual formulations take the normativity of (true) belief to be concerned with the relation between an attitude and the truth of a given proposition. However, the normativity of content is concerned with statements of content-individuating norms for a huge variety of mental contents. It is not even remotely plausible that these different projects turn out to be the same. What this suggests is that no satisfactory assessment of the normativity of mental content will be in view until we do not pay sufficient attention to the full explanatory power of IRS theories.

Acknowledgements:

This research has received financial support from the Ministerio de Economía y Competitividad, Spanish government, via the research project FFI2009-08828/FISO, and from the Catalan government, under the consolidated research group GRECC, SGR2009-1528.

REFERENCES

Bykvist, K. and A. Hattiangadi 2007, 'Does thought imply ought?', *Analysis* 67, 277-85.
Boghossian, P.A. 2003, 'The normativity of content', *Philosophical Issues* 13, 31-45.
Boghossian, P.A. 2005, 'Is meaning normative?', in Beckermann, A. and Nimtz, C. (eds.), *Philosophy – Science – Scientific Philosophy*, Paderborn, Mentis.
Broome, J. 1999, 'Normative requirements', *Ratio* 12, 398-419.
Gibbard, A. 2005, 'Truth and correct belief', *Philosophical Issues* 15, 338-50.
Glüer K. and A. Wikforss. 2009, 'Against content normativity', *Mind* 118, 31-70.
Hattiangadi, A. 2006, 'Is meaning normative?', *Mind and Language* 21, 220-40.
Hattiangadi, A. 2009, 'Some more thoughts on semantic oughts: a reply to Daniel Whiting,' *Analysis* 69, 54-63.
Jarvis, W.B. 2012, 'Norms of intentionality: norms that don't guide', *Philosophical Studies* 157, 1-25.
Kripke, S. 1982, *Wittgenstein on Rules and Private Language*. Oxford, Blackwell.
Miller, A. 2008, 'Thoughts, oughts and the conceptual primacy of belief', *Analysis* 68, 234-38.
Peacocke, C. 1992, *A Study of Concepts*. Oxford, Oxford University Press.
Verdejo, V.M. 2009a, *The Rational and Natural Mind.* Saarbrücken: VDM Verlag.
Verdejo, V.M. 2009b, 'Why rationalist compositionality won't go away (either)', *Theoria* 24, 29-47.

Verdejo, V.M. forthcoming, 'The rationalist reply to Fodor's analyticity and circularity challenge, *Theoria*.
Wedgwood, R. 2002, 'The aim of belief', *Philosophical Perspectives* 16, 276-97.
Wedgwood, R. 2007, *The Nature of Normativity*. Oxford, Oxford University Press.

EL ROMPECABEZAS DE LA ATRIBUCIÓN DE CREENCIAS

IGNACIO VICARIO
UAM

ABSTRACT: David Sosa (1996) thinks that Kripke's defence of Millianism against the traditional argument based on belief reports fails because it presupposes principles that are justified only by Millianism. I criticize some points in Sosa's discussion. And I propose that both the traditional argument and Kripke's analogous arguments presuppose a relational principle of belief reports. That principle is not implied by Millianism, but, importantly, it falls short of explicating the typical context-sensitive of belief reports. So I suggest this may be the source of the puzzling air of the argumen**ts.**
KEYWORDS: ambigüedad, millianismo, principio relacional, sensibilidad contextual

I. En "A Puzzle About Belief" (1979), Saul Kripke planteó un rompecabezas sobre la creencia con el que trató de responder a un argumento tradicional fregeano, basado en las oraciones de creencia, en contra de la teoría milliana de los nombres propios; y, por ende, quiso mostrar que las oraciones de atribución de creencia no eran terreno adecuado para sacar conclusiones sobre la semántica de los nombres propios. Ambos objetivos quedarían minados si el planteamiento del rompecabezas de Kripke presupusiera de algún modo una semántica milliana. Precisamente, ésta es una crítica que con frecuencia se le ha hecho. Mi propósito es defender las conclusiones de Kripke de una de estas críticas, elaborada por David Sosa en "The Import of the Puzzle About Belief" (1996).[1] Apuntaré que el rompecabezas deriva de otro principio que supone un análisis relacional de las oraciones de creencia.

II. El Millianismo (M) es una teoría sobre la semántica de los nombres propios que afirma que su significado se agota en la referencia. (M) parece implicar la validez, en general, del principio de substitución, (S), por el que términos correferenciales son intercambiables salva veritate.

Desde posiciones antimillianas, se ha argüido en contra de (M) sobre la base de las oraciones de atribución de creencia. El argumento tradicionalmente utilizado es éste: Supongamos que Jones es racional, y que asiente a "Cicerón era calvo" y "Tulio no era calvo". Ordinariamente,

[1] Taschek (1988) ofrece otra crítica en la línea de que el rompecabezas se plantea por la admisión de principios que el millianismo admite, pero no el fregeanismo.

aplicaríamos el principio (D): Si un hablante normal del castellano, meditada y sinceramente, asiente a "p", entonces cree que p, y obtendríamos: Jones cree que Cicerón era calvo y Jones cree que Tulio no era calvo. Dado esto último, por (S), se sigue que Jones cree que Cicerón no era calvo. Pero: (*) Si Jones cree que Cicerón era calvo y Jones cree que Cicerón no era calvo, cree una contradicción. Pero un agente racional no cree una contradicción.[2] Luego, Jones no es racional.[3] La responsabilidad del absurdo se imputa a (S), refutando con ello (M), que lo implica.

Kripke (1979) disiente de esta conclusión. El absurdo puede reproducirse sin (S), y sin (M). La propia práctica de atribuir creencias es problemática, genera un rompecabezas. Su ilustración del rompecabezas más popular se basa en el ejemplo de Pierre y sus creencias sobre Londres. Por simplicidad, me centraré en el otro: Peter sabe de un famoso pianista llamado "Paderewski", lo que le lleva a asentir a "Paderewski tiene talento musical". Después tiene noticia del Primer Ministro polaco Paderewski, y desconfiando de las habilidades musicales de los políticos, asiente a "Paderewski no tiene talento musical". Puesto que Paderewski, el pianista, es en realidad el Primer Ministro, la situación descrita genera un argumento análogo al tradicional, salvo por la importante diferencia de que ahora no se utiliza (S). Si la misma situación paradójica puede producirse sin utilizar (S), es que (S) no era la causa, bloqueando el argumento tradicional contra (M).

III. Distanciándose de algunas discusiones del rompecabezas que cuestionan la aceptabilidad de los principios utilizados explícitamente por Kripke, como (D),[4] Sosa juzga (y yo coincido) que el rompecabezas no depende sustancialmente de su utilización, por cuanto puede recrearse sin ellos (Sosa 1996, 383-85). La metodología argumentativa seguida por Kripke es correcta, por tanto, a fin de mostrar que el argumento tradicional no consigue minar (S). Sin embargo, Sosa considera que, en último término, Kripke no consigue con sus argumentos análogos defender (M) del argumento tradicional, por cuanto se requiere para formularlos de otro principio, distinto de (S), pero igualmente implicado por (M). Es decir, la idea es que de una manera o de otra, el rompecabezas surgiría únicamente por algún principio, ya sea (S) u otro, implicado por (M), pero no por el fregeanismo. Con lo que, a la postre, (M) provoca el rompecabezas de la creencia (Sosa 1996, 393).

La cuestión clave es qué justifica en el análogo de Kripke este paso (**): "Si Peter cree que Paderewski tiene talento musical y Peter cree que Paderewski no tiene talento musical, entonces

[2] No pocos millianos rechazan este punto, que dejaré de lado en la discusión.

[3] Sigo una formulación próxima a la dada por Sosa (1996), más detallada que la de Kripke. La explicitación de la utilización de (D) se debe a Kripke.

[4] Para una crítica a (D), véase Moore (1999). Por otra parte, Taschek (1988) cuestiona el principio de traducción utilizado en el famoso ejemplo de Pierre.

Peter tiene creencias contradictorias". Un elemento justificador es el supuesto siguiente: Un sujeto tiene creencias contradictorias si los contenidos de sus creencias son contradictorios (Sosa 1996, 388). Pero hay algo más involucrado, según Sosa, algo que afecta a la representación lógica de esos contenidos. Téngase en cuenta que los nombres propios de las lenguas ordinarias pueden ser ambiguos. Si las oraciones en cuestión fueran "París es bonita" y "París no es bonita", donde en la primera se hace referencia a la capital de Francia y en la segunda, a la población de Texas (EE. UU.), naturalmente no se expresarían contenidos contradictorios. Y un sujeto podría formar creencias con esos contenidos y no tendría creencias contradictorias. ¿Qué separa, pues –se pregunta Sosa (1996, 387)–, la situación enunciada en (**) de una situación en la que un sujeto, Rock, cree que París es bonita y cree también que París no es bonita (en las interpretaciones mencionadas)? La diferencia es clara, nos dice Sosa, en el paso (**) "Paderewski" tiene las dos veces la misma referencia, pero no "París" en la afirmación sobre las creencias de Rock. De ello concluye Sosa (1996, 388) que, en la justificación del paso de (**) (y (*), en el tradicional), el principio utilizado para la representación lógica de los contenidos de las oraciones es (H): "Si un nombre del lenguaje ordinario tiene solamente un *referente* entonces puede, de forma correcta, representarse lógicamente por una única constante".

Pero "si (H) es *esencial* en recrear los resultados contradictorios [...], entonces parecería que [tanto el argumento tradicional como los análogos de Kripke] son reducciones al absurdo de (H)" (Sosa 1996, 388-89, el subrayado es mío). Mas aquí no acaba la cosa. Dado que el principio establece la correferencia como garante de que dos expresiones de nombre propio no son ambiguas, este principio es implicado por (M); pero no por el fregeanismo. Éste "postula un tipo de significado que determina, pero que no es determinado por el tipo de significado postulado por el millianismo" (Sosa 1996, 390). Así: "*El fregeanismo puede ver* ambigüedades *para las que el millianismo está ciego*" (Sosa 1996, 389). Por ejemplo, "Cicerón" y "Tulio" serían ambiguos para el primero, no para el segundo. Y Sosa considera que también es verosímil que "Paderewski" sea ambiguo en el ejemplo.[5]

Desde el punto de vista de Sosa, los pasos (*) y (**) no estarían debidamente justificados para el fregeano.[6] Y la lección sería clara: "Mi sugerencia es que un paso *esencial* de los [argumentos] análogos [de Kripke] es justificado apelando a un principio, (H), que a su vez está justificado sólo por el millianismo. Los ejemplos entonces equivalen a otras tantas reducciones al absurdo del principio y con éste, del millianismo" (Sosa 1996, 389, el subrayado es mío). Según

[5] O "Londres", en el otro ejemplo de Kripke. Véase Sosa (1996), 389 y 394.

[6] Si "Paderewski" es ambiguo, cabe decir tanto que el paso (**) no está justificado, como que no debería haberse aplicado, de antemano, (D). Véase Sosa (1996), 390 n. 13 y 394.

esto, Kripke no habría logrado desligar al millianismo del problema puesto de manifiesto por el argumento tradicional. Tanto en éste como en los argumentos análogos ofrecidos por Kripke, se estaría presuponiendo una semántica de los nombres propios milliana, que sería la fuente del problema.

IV. Discrepo de que la razón del rompecabezas sea la presuposición de una u otra semántica de los nombres propios. Antes bien creo que su origen está en la propia práctica de atribuir creencias –como sugería Kripke–, en las oraciones con las que informamos de las creencias de los sujetos. Eso es lo que pretendo motivar.

Pero empecemos con (H). No veo que sea *esencial* para producir el rompecabezas. Suponer que (H) es lo que permite distinguir entre el caso de las creencias de Rock y Peter, inclina innecesariamente la cuestión en contra del millianismo. El principio podría formularse de un modo más neutro, sin presuponer un tipo de valor semántico o constituyente proposicional. Bastaría, simplemente, con cambiar "referencia" por "valor semántico" en la formulación dada por Sosa. Obtendríamos (H'): "Si un nombre del lenguaje ordinario tiene solamente un *valor semántico* entonces puede, de forma correcta, representarse lógicamente por una única constante". Para que dos expresiones de un nombre propio sean igualmente representadas en el contenido basta que compartan el valor semántico, ya sea éste la referencia, ya sea un sentido que determine la referencia (u otra opción). Formulado así el principio no presupone el millianismo, y puede ser aceptado por un fregeano (como el propio Frege). Y la defensa de Kripke podría seguir en pie.

Es cierto que muchos neo-fregeanos querrían replicar que la opción fregeana preferible sostiene que el sentido de un nombre propio puede cambiar de boca en boca, y aun de uso en uso. Sosa parece admitir tal cosa.

Si se acepta pues que "Paderewski" es ambiguo en boca de Peter, que cambia su sentido (en el ejemplo) de una ocasión a otra, entonces los ejemplos de Kripke no parecerían suponer un rompecabezas para el neo-fregeano, porque (H') no se aplicaría en ese caso. Pero fijémonos que entonces, la discusión ya no estaría centrada estrictamente en si Kripke *presupone* un principio milliano, sino en si el valor semántico (sea cual sea) de los nombres propios es estable e intersubjetivo, o si es sensible al contexto.[7] Naturalmente, esta otra cuestión es controvertida. Kripke (1979, 246, y 259-61) da argumentos independientes para descartar esa salida al fregeano, y Sosa (1996, 396-99) le replica. No quiero detenerme en este frente de la vieja batalla entre

[7] Por *sensibilidad contextual* de los nombres, no incluyo la dependencia *pre-semántica* vista en el ejemplo de "París". La falta de espacio me impide ser más preciso con estas diferencias. Dejo al lector la inteligencia del sentido en discusión.

millianismo y fregeanismo, pero es claro que el neo-fregeano tiene una larga tarea por delante, si quiere no sólo plantear una posibilidad teórica a fin de zafarse del rompecabezas, sino darle la debida consistencia.[8]

Volvamos a la cuestión de la neutralidad del rompecabezas planteado por Kripke, y si puede detener el argumento antimilliano basado en las oraciones de creencia. El fregeano (ahí está ese argumento tradicional y sus antecedentes en Frege) ha solido considerar que las oraciones de creencia eran un terreno más (uno favorable, eso sí) donde dirimir esa batalla sobre el significado de los nombres propios. Sosa, al hacer recaer la discusión sobre (H), también ha privilegiado esa aproximación. Kripke, por el contrario, al tratar de disculpar al millianismo del argumento tradicional, sugiere que, más bien, esa discusión debería alejarse de ese terreno, que las oraciones de creencia, o la práctica que sustenta su uso, son un caso especial, uno difícil, que presenta sus propios problemas. Creo que no le falta razón.

En la discusión de Sosa (como en otras que tratan de sostener que Kripke presupone el millianismo), no se destaca otro elemento que estaría operando en los pasos (*) y (**); un elemento relativo a la semántica de las propias oraciones de creencia. Implícitamente, se está aceptando un modo de analizar estas oraciones complejas. Puede sorprender que ocupados precisamente con ellas, se pase por alto.[9] Probablemente, ha sido interiorizado, y se supone por defecto: el análisis es sencillo, y tradicionalmente ha gozado de una amplia aceptación.[10] Es el análisis relacional de las oraciones de creencia. Según el cual, se entiende que las creencias son relaciones entre sujetos y contenidos o proposiciones; que las oraciones de creencia reflejarían esa relación, y que, característicamente, la cláusula subordinada funcionaría como un término referencial que especifica el contenido creído. El análisis puede formularse así, (R): "*S* cree que *p*" establece un relación de creencia entre *S* y el contenido denotado por la cláusula "que *p*", siendo el contenido de esta cláusula la proposición expresada por "*p*".[11]

[8] Hace falta algo más que juzgar que "no es verosímil suponer que Peter puede asentir racionalmente a 'Paderewski tiene talento musical' y a 'Paderewski no tiene talento musical' a menos que asocie diferentes sentidos con 'Paderewski'" (Sosa 1996, 394). ¿Es mi nombre ambiguo porque un individuo no me re-identifica a través de él en una ocasión? Hay que dar consistencia teórica a la noción idiosincrásica de sentido apelada ahí para justificar esa diferencia de actitud (que puede reproducirse también con términos generales), de modo que pueda ser el valor semántico de la expresión, y que con ello se pueda cimentar que dos personas lleguen a tener creencias contradictorias. Es más, debe explicarse adecuadamente cómo es que, aun suponiendo que las creencias de Peter sobre Paderewski no sean contradictorias, sí parecerían claramente contradictorias en su boca las afirmaciones "Paderewski tiene talento musical" y "Paderewski no tiene talento musical". En Vicario (2002) y (2009) he criticado algunos aspectos de que el significado de un nombre propio varíe contextualmente.

[9] Kripke (1979, 242) no lo pasa por alto.

[10] Entre fregeanos y millianos.

[11] Véase Récanati (1993), 349-51.

Este análisis no presupone un modelo particular de proposición (fregeano o milliano). Pero sí hace coincidir de modo estable el contenido de la proposición de la oración subordinada y el contenido creído que se atribuye al sujeto. De suerte que si dos oraciones simples "p1" y "p2" expresan una misma proposición, entonces al sujeto se le atribuye el mismo contenido, y "S cree que p1" y "S cree que p2" expresarán la misma proposición (y tendrán las mismas condiciones de verdad). Ahora bien, si algo sabemos sobre estas oraciones es la altísima sensibilidad contextual que presentan los juicios de los hablantes sobre los valores de verdad que les corresponden. La misma oración de creencia, que involucre al mismo hablante y a los mismos sujetos y objetos, parece que puede dar lugar a una atribución correcta o incorrecta según el contexto.[12] ¿A qué se debe esta sensibilidad contextual de las oraciones de creencia (de la práctica de atribución)? La línea de Sosa fijaría la causa de esa variabilidad contextual en los valores semánticos de los nombres propios. Pero existe otra posibilidad altamente verosímil: la flexibilidad en la relación entre el contenido atribuido y la proposición correspondiente a la oración subordinada, de suerte que aquél pueda ir más allá de ésta; flexibilidad que podría ser desencadenada contextualmente por algún factor de la estructura de las oraciones (ya sea la cláusula subordinada, el verbo, un constituyente inarticulado...).[13] Esta posibilidad, que es compatible con (M), entra en conflicto, de un modo u otro, con (R).[14]

Volvamos al argumento tradicional y los basados en los ejemplos de Kripke. Todos ellos derivan en un absurdo. ¿Qué está mal? El antimilliano tradicional utiliza el primer argumento para cuestionar (S) y con ello (M). Kripke muestra que el problema se reproduce sin (S), con lo que esa no debe ser la causa (y por tanto no hay que suponer que la causa es (M)). Sosa está de acuerdo con la exoneración de Kripke de (S). Pero no, con que haya librado a (M), porque en todos los argumentos se utilizaría (H) que es implicado por (M). Ya he corregido que igualmente se podía haber utilizado (H'), que no presupone (M). Dejado ahí el asunto, Kripke no presupondría (M), pero no evitaría que alguna versión neo-fregeana estuviera en ventaja al bloquear la derivación del rompecabezas (la aplicación de (H')), si se acepta que los valores semánticos de los nombres varían contextualmente.

Mi sugerencia es que (R) es utilizado en todos esos argumentos. Y que muy bien la fuente del rompecabezas pueda ser precisamente (R), y no (M) o algún principio implicado por (M). El

[12] Sobre la sensibilidad contextual de las oraciones de creencia, véase, por ejemplo, Richard (1990) o Récanati (1993).

[13] Otra opción es que la función de las atribuciones de creencia no sea la de especificar el *contenido* de la creencia. Véase Bach (1997). Opción que también choca con (R).

[14] Una discrepancia con (R) (con el modo de la oración de determinar del contenido) no supone necesariamente una discrepancia con el carácter relacional de las propias creencias. Por otro lado, ejemplos de la compatibilidad con (M) serían análisis como el de Richard (1990) o Crimmins y Perry (1989).

fenómeno de la atribución de creencia y por tanto de las oraciones de creencias exhibe una fuerte dependencia contextual. El punto crucial es que esa dependencia no es recogida por *ninguno* de los principios utilizados en la derivación del rompecabezas. Pero, en sí mismo, el rompecabezas no permite decir si la derivación del absurdo se produce por un principio que no tiene en cuenta una posible sensibilidad contextual de los nombres propios (sea (H') u otro) o por un principio que no tiene en cuenta la sensibilidad contextual del mecanismo por el que la oración de creencia determina el contenido atribuido (como (R)).[15] Como esta última posibilidad es perfectamente compatible con que los nombres sean millianos (incluso en estos contextos), Kripke sí tenía razón en que el argumento antimilliano, por sí solo, no permite concluir la falsedad de (M); y en que el fenómeno de la atribución de creencia no es el mejor lugar para dirimir la semántica de los nombres propios. "Los casos difíciles producen malas leyes" (Kripke 1979, 270).[16]

REFERENCIAS

Bach, K. 1997, 'Do Belief Reports Report Beliefs', *Pacific Philosophical Quarterly* 78 (3), 215-241.

Crimmins, M. y Perry, J. 1989, 'The Prince and the Phone Booth: Reporting Puzzling Beliefs', *The Journal of Philosophy* 86 (12), 685-711.

Kripke, S. A. 1979, 'A puzzle about belief', en Margalit, A. (comp.) *Meaning and Use*, Dordrecht & Boston, Reidel, 239-283.

Lewis, D. 1981, 'What puzzling Pierre does not believe', *Australasian Journal of Philosophy*, 59 (3), 283-289.

Moore, J. G. 1999, 'Misdisquotation and substitutivity: when not to infer belief from assent', *Mind* 108 (430), 335-366.

Récanati, F. 1993, *Direct reference : from language to thought*, Oxford, Blackwell.

Richard, M. 1990, *Propositional attitudes: an essay on thoughts and how we ascribe them*, Cambridge: Cambridge University Press.

Sosa, D. 1996, 'The Import of the Puzzle About Belief' *The Philosophical Review*, 105 (3), 373-402.

Taschek, W. 1988, 'Would a Fregean be Puzzled by Pierre?', *Mind* 97 (385), 99-104.

Vicario, I. 2002, '"Paderewski" y el problema del valor cognoscitivo en Frege', *Revista de Filosofía* 27 (2), 361-387.

Vicario, I. 2009, 'Insuficiencias del modelo deíctico de los nombres propios', *Crítica* 41 (123), 95-127.

[15] En Lewis (1981) se ofrece una lectura próxima del rompecabezas, en cuanto se indica que lo rechazado es un cuadro teórico sobre la atribución de creencia, más que una tesis particular de éste.

[16] Diversas instituciones públicas han contribuido a la financiación de las investigaciones de las que este trabajo es deudor: Programa CONSOLIDER-INGENIO 2010, "Perspectival Thoughts and Facts" (CSD2009-00056), MICINN (Gobierno de España); proyecto "Conocimiento, referencia y realismo" (FFI2011-29560-C02-01), MICINN; proyecto "Significado, Traducción y Contexto" (FFI2009-11347), MICINN; y grupo de investigación consolidado LOGOS (2009 SGR 1077), AGAUR (Generalitat de Catalunya).

LA PERCEPCIÓN DEL ASPECTO EN LAS INVESTIGACIONES FILOSÓFICAS

JAVIER VILANOVA ARIAS
Universidad Complutense de Madrid
vilanova@filos.ucm.es

ABSTRACT: The paper provides an interpretation for Wittgenstein´s analysis of the notion of "see as". The paper argues that the main point for a correct interpretation is to understand similarities and differences between things in a realistic way, as objective, in re aspects of perception.
KEYWORDS: Wittgenstein, see as, perception.

1. El problema

El tema de este trabajo es una familia de nociones emparentadas por el vínculo entre percepción y pensamiento que se da en ellas, y entre las cuales resulta central la nociones de "ver como" y "percepción del aspecto". Es evidente el valor que tiene el tema para la psicología, la filosofía de la percepción y la filosofía de la mente en general. También tiene vínculos con la Estética y la Filosofía del Arte, así como la propia Filosofía de la Filosofía. Pero parece que el interés de Wittgenstein está más conectado con la filosofía del lenguaje: es como si buscara en él la clave para entender la «vivencia del significado» así como el tipo de destreza (¿conocimiento?) asociado a la competencia lingüística. A pesar de ello, tanto el tratamiento que hace Wittgenstein como su interpretación son esquivos y propensos a la confusión. Para empezar, sería engañoso decir que Wittgenstein pretende dar un análisis (y mucho menos una definición o una teoría filosófica) de tales nociones. Eso iría en contra de su concepción "terapéutica" de la Filosofía. En realidad Wittgenstein pretende aclarar el uso de una serie de figuras, y disolver algunas confusiones que engendra su uso. Witgenstein presenta (como es habitual en él) una gran cantidad de casos muy variados (no hay que esperar que las conclusiones se apliquen a todos): la caja de zapatos, la cabeza C-P, el triángulo, la cruz blanca-negra, el rostro sonriente, el cubo de Neckler, el peldaño convexo, la fila de puntos…. Además, hay una serie de casos relacionados, cuyos parecidos y diferencias con los casos de estudio explota: reconocer a un viejo conocido, o a un objeto familiar en una posición extraña, ver el dibujo de un caballo en un cuadro como galopando, o el dibujo de una esfera como flotando, reconocer un pasaje musical como una «marcha» o una «danza», percibir la «atmósfera» de una palabra.

Se podría decir que el propósito clarificador de Wittgenstein se dirige a la siguiente familia de nociones, cuyo uso debe se aclarado:

(a) Ver A como B.
(b) Ver el aspecto x de A.
(c) cambio de aspecto.
(d) Ceguera a un aspecto.

Principalmente, hay que disolver una posible fuente de confusiones que mana de la naturaleza aparentemente paradójica del «cambio de aspecto»: en efecto, ante una "doble figura" como cabeza C-P, al vivir el cambio de aspecto, informamos de que seguimos viendo lo mismo pero sin embargo vemos algo distinto. Pero esto parece una contradicción[1]. Ligada a esta confusión aparece otra, no menor. Y es que el "ver como" parece surgir de una determinada interpretación de la figura por parte del sujeto. Pero una interpretación es entendida primariamente como un fenómeno conceptual, ligado a eventos mentales e incluso al juicio. Entonces, ¿cómo es posible que vea una cosa de acuerdo con una interpretación?

2. Una explicación equivocada

Hay un tipo de explicaciones del cambio de aspecto, en clave constructivista, que el mismo Wittgenstein se encarga de rechazar explícitamente. Según estas explicaciones, el cambio de aspecto es producido por la manera en que el espectador organiza el material "bruto" de la sensación añadiendo estructura, forma o configuración. De esta forma pasan a hablar de dos entidades distintas: la sensación "bruta" (la percepción no organizada por el sujeto) y la imagen subjetiva o "interna" (la percepción ya organizada). Según estas explicaciones, lo que varia en el cambio de aspecto son elementos o propiedades de la imagen subjetiva o «interna» (un objeto mental).

Así es cómo lo explica la Psicología de la Gestalt, al entender el cambio de aspecto como variaciones de la «organización» de la forma y el color que constituyen el *input* perceptivo. Pero esto, como Wittgenstein alega, contradice el informe que damos en primera persona de la experiencia: seguimos viendo lo mismo (nada cambia en mi sensación; de hecho, no soy capaz de

[1] 474. Lo que resulta incomprensible es que nada haya cambiado y que, no obstante, todo lo haya hecho. Ésta es, en realidad, la única manera de formularlo. Esta otra forma de hacerlo es incorrecta: en cierto sentido no ha cambiado pero sí lo ha hecho en otro.(BPP, vol. 2, pág. 86c)

describir nada nuevo en mi sensación salvo el informe de primera persona «ahora lo veo como…»). Estas explicaciones dan lugar a la siguiente:

Paradoja de la imagen interna: no «vemos» que cambie nada en nuestra experiencia visual cuando cambia el aspecto.

De ahí que Wittgenstein deseche la idea de «figura interna» como un mito filosófico[2]. En base a estas mismas razones por las que Wittgenstein rechaza algunos análisis del cambio de aspecto podemos rechazar nosotros algunas interpretaciones de su propio análisis (cf. Peacocke 1983 y Mulhall 1990 y 2001), incompatibles tanto con el anti-mentalismo como con el anti-representacionalismo de las Investigaciones.

3. Una exégesis acertada.

La exégesis adecuada, según lo que vengo defiendo, es la que explica el cambio de aspecto como un cambio en la interpretación. Y aquí «interpretación» ha ser entendido como una acción cognitiva, como un evento o acontecimiento mental similar a otros como juzgar o entender, y, por lo tanto:

- tiene naturaleza conceptual, es del tipo de «pensar»
- requiere aprendizaje,
- es un acto,
- puede ser correcto o incorrecto (verdadero o falso).
- es, o puede ser, consciente,
- es, o puede ser, hecho voluntariamente[3].

Aunque considero que la exégesis correcta pasa, como hemos dicho, por entender el «ver como» como el ver según una interpretación (incluso cabría decir, aquí, según un concepto, aunque, en el contexto de las Investigaciones, es más correcto decir «según un aire de familia») y el cambio de aspecto como un cambio en la interpretación (manteniéndose indemne la sensación), ésta no está exenta de problemas. Dejando aparte problemas puramente exegéticos, sobre los que diré algo más adelante, la explicación propuesta por Wittgenstein deja aspectos de la experiencia cotidiana del

[2] PU, pág. 451.

[3] En esta línea se inscriben, acertadamente según mi punto de vista, Budd 1987 y Schroeder 2010:

"We have seen that the intrinsic nature of seeing something in accordance with an interpretation does not have an essence that is capable of being specified independently of the interpretation" (Budd 1987, pág. 13).

To conclude: visual aspect perception may well be called 'seeing', although it is often more concept-laden than seeing just shapes and colours. An important result of the discussion so far is that one's visual impression cannot be construed as an inner picture. (Schroeder 2010, pág. 363)

«ver como» sin explicar, y, lo que resulta más grave, parece ser incompatible con otros). Todos ellos tienen que ver con esa naturaleza "mixta" del ver como (mitad percibir, mitad pensar), que hace que aparezcan propiedades del ver como que no encajan (o no parecen encajar) en una u otra familia de vivencias (familia de la percepción o familia del pensamiento). Esto lo podemos comprobar fácilmente haciendo un repaso a la lista que hace Schroeder 2010 (los dos primeros problemas son señalados por Budd 1987):

(a) interpretar es una acción, el ver como es continuo,
(b) interpretar puede ser falso o incorrecto, el ver como no parece serlo,
(c) no podemos describir la «visión sin interpretación»,
(d) reaccionamos frente al aspecto como si lo percibiéramos,
(e) ver como es ver un parecido.

De los cinco problemas señalados por Schroeder, los dos primeros son los fundamentales. El quinto, como se verá más tarde, puede no ser tanto un problema como una parte importante de la solución. El tercero y el cuarto permiten una respuesta rápida:

(c) No es cierto que no podamos describir la visión sin interpretación. En muchos casos esto es muy sencillo (veo un cuadrado, y otro cuadrado más arriba y a la derecha que se solapa con el anterior, y una serie de líneas diagonales que unen los vértices de los cuadrados). En otros casos esto se logra hacer después de un entrenamiento adecuado (un pintor, un detective, un notario…). Además, que elijamos la descripción del aspecto como un modo de describir la percepción es natural, ya que es un recurso rápido y eficiente, pero NO SIEMPRE lo elegiríamos (¿una casa para el cubo de Neckler?)

(d) Es natural que reaccionemos ante la figura como ante un objeto: forma parte de nuestra forma de vida como seres humanos que reaccionemos ante representaciones como ante objetos reales. Esto mismo ocurre con las palabras. Los ejemplos se pueden multiplicar aquí: la señal de stop no me dejó pasar; Juan me obligó (cuando Juan sólo emitió la oración «hazlo»); en mi casa tengo elefantes, jirafas, hipopótamos... Precisamente uno de los temas que Wittgenstein también discute en las UP es la experiencia de los signos y figuras como algo «vivo»[4]

[4] PU, pág. 461, PU, pág. 471

4. La"ontología" de las Investigaciones

Los problemas (1) y (2) remiten a la misma cuestión: lo que no acaba de dejar claro Wittgenstein es si el aspecto es algo "visto" o es algo "pensado". El problema estriba en que si es algo "pensado" entonces, una vez más, es algo puesto por el sujeto y volvemos a los problemas de 3 y 4. Pero si es algo "visto", no se explica satisfactoriamente que pueda ser aprendido (requiere técnica) ni que haya personas "ciegas al aspecto". Schroeder (p. 362) y Budd (p.17) advierten tal dilema interpretativo, pero, en mi opinión, no lo resuelven adecuadamente mediante el recurso a la naturaleza "híbrida" (entre interpretación y percepción) del ver cómo. Es cierto que, como ocurre continuamente en las IF, lo que se presenta es una familia de casos con diferencias notables entre sí y que no necesariamente han de poseer una propiedad común (hay casos en que el aspecto es más "visto" y otros en que es más "interpretado"), y también es cierto que el propio Wittgenstein entiende el ver-cómo como un intermedio entre percibir y pensar. Pero esta naturaleza híbrida es, como defenderé, precisamente lo que hay que explicar.

En mi opinión, este dilema interpretativo sólo se puede resolver internándose en un tema que, con respecto a las Investigaciones, resulta bastante espinoso, y unánimemente evitado por parte de los exégetas: ¿cuál es el status ontológico de los aires de familia? La cual, a su vez, nos remite a esta otra: ¿tienen los parecidos y diferencias entre cosas, a partir de las cuales se construyen los aires de familia, un fundamento *in re*?

5. Resolviendo problemas.

Aunque he venido usando el término «conceptual» para describir la naturaleza de la interpretación, la palabra «concepto» puede resultar engañosa si se utiliza de la manera tradicional. Para Wittgenstein no existen conceptos, si estos se entienden como entidades abstractas, como formas sin materia o como objetos mentales (abstracto). Tampoco existen conceptos si se toman como entidades cerradas y precisas que determinan unívocamente una extensión (cerrado). Las Investigaciones están plagadas de argumentos contra estas nociones (sobre seguir una regla, indeterminación de la definición ostensiva, lenguaje privado, preeminencia del contexto, etc…) Para sustituir esta noción de concepto abstracto cerrado Wittgenstein propone la de «aire de familia»: parecidos y diferencias entre una serie de cosas. Ahora bien, aun cuando la asociación de una palabra a un aire de familia es fruto de una convención, y por lo tanto algo voluble y no objetivo (y mucho más cuando, como es el caso, la convención misma es cambiante e imprecisa) el aire de familia en sí no tiene que serlo.

Según mi lectura de las Investigaciones, los aires de familia son OBJETIVOS: los parecidos entre los objetos no los pone el hablante o los hablantes (esto sería caer en el mentalismo contra el que las Investigaciones luchan tan encarnizadamente), sino que están ya en el mundo, en las cosas mismas. Otra cosa es que, dada la gran cantidad de aspectos en que se parecen o se diferencian las cosas (recordemos que dijimos que eran no definibles algebraicamente), y teniendo en cuenta que las convenciones lingüísticas se fijan comunitariamente, haya siempre un umbral de incertidumbre sobre cuál es el o los aires de familia asociados a una palabra y como deben ser utilizados en el contexto de uso particular.

Por ello hay siempre una indeterminación sobre el uso de una palabra en un contexto: la aplicación en los casos pasados no determinan unívocamente la aplicación en el presente, pues hay distintas analogías y diferencias que puedo tener en cuenta; y tampoco puedo acudir a rasgos formales o nociones abstractas, pues no hay nada fuera de los parecidos que los fundamente.

Ahora bien, en los casos felices, en los que alguien usa una palabra de la manera adecuada para la situación, se produce una INTERPRETACIÓN correcta: asociamos la palabra con los parecidos y diferencias relevantes y adecuados. La interpretación es un acto de naturaleza eminentemente lingüística (exige manejar una convención), pero la acción asociada de detectar los parecidos es no lingüística: de hecho, en la medida que los parecidos son propiedades objetivas de las cosas, ocurre que los «percibimos» (no los pensamos, ni los concebimos, ni los imaginamos, ni los decidimos convencionalmente). Y ello no se contradice con el hecho de que sea necesaria la adquisición de una destreza o habilidad lingüística para poder detectar ciertos parecidos y utilizarlos adecuadamente.

En resumen: percibir los parecidos y diferencias es tanto una parte del aprendizaje como un acto colateral al uso del lenguaje.

Ahora es posible entender mejor por qué dice Wittgenstein que interpretar es algo parecido a ver, o, mejor dicho, es «a medias» una vivencia visual (y a medias un pensamiento). Y por ello, además, es tan importante el análisis de estas nociones para Wittgenstein: porque al poner de manifiesto la objetividad del aspecto (es visto), revela, de paso, la objetividad de la interpretación lingüística, en particular, y del pensamiento humano en general. También este punto sirve para explicar el porqué dela «gradación» de casos proporcionados. En los casos más simples (cruz negra-blanca p. ej.) es más fácil en fundamento in re de la interpretación. Que Wittgenstein nos convenza de la existencia de un aire de familia con los casos más complicados (una proyección geométrica, p. ej.) no sólo nos hará ver a su vez el fundamento *in re* de estos últimos, sino que además sirve para

explicar cómo se aprenden y/o como se generan las lenguas (otra de las preocupaciones de Wittgenstein desde el principio del libro)[5].

También entendemos mejor porque el «ver como» parece experimentarse, en muchos casos, como continuo. Porque aunque el cambio de aspecto es el producto de una interpretación (que sí es un acto) el apercibimiento de los parecidos no es a su vez una interpretación. Ahora bien, en la mayoría de los casos la percepción involucrada en el «ver como» es atípica, pues supone ver la analogía de una figura presente con otros no presentes (involucra a la memoria y a la imaginación). Esto explica algunos de sus rasgos peculiares, así como el hecho de que requiera en muchas ocasiones la adquisición de destrezas especiales (es «sofisticada»). Pero la intervención de la memoria y la imaginación, así como el requisito de destrezas previas:

(i) no impide que podamos usar con propiedad la palabra percepción (¿puedes ver que esta cartulina tiene el mismo color que la que te enseñé hace un momento?)

(ii) no es un rasgo esencial del «ver como» (como el mismo Wittgenstein dice, podemos colocar la cabeza pato entre otras figuras de conejo, o entre otras figuras depato, y prescindir del recurso a la memoria; o podemos ver de forma espontánea la cruz negra o blanca, sin poseer ni siquiera lenguaje).[6]

También podemos argumentar ahora que, contrariamente a lo que dice Schroeder, uno puede «ver como» falsa o incorrectamente. Primero, porque uno puede creer percibir analogías objetivas que en realidad no existen (ilusiones ópticas al nivel de los parecidos y diferencias). Esto es lo que se daría en situaciones en las que hay alguna anomalía en nuestro entorno o en nuestro sistema perceptivo, o simplemente por descuido (errores del lado de la percepción). Segundo, porque uno puede identificar un aspecto que no es el apropiado en la situación de habla y, consecuentemente, hacer una mala interpretación. Así, cuando a uno se le invita a ver un aspa en la figura de la cruz blanca-negra puede verla como el aspa de un avión (líneas negras) cuando en realidad se desea que vea el aspa de un molino (líneas blancas), y esto puede tener consecuencias negativas en la situación de habla (por ejemplo, si te pido que me traigas «un aspa como ésta».

Ahora también estamos en disposición de entender mejor el punto que Schroeder entendía como problemático y que para mí es parte de la solución: explicar en que sentido ver como es ver un parecido.

5 BPP, vol. 2t pág. 93c.
6 WLPP, pág. 147, P.U. , pág. 479.

Ya hemos dicho que el aspecto es (salvo en los casos más simples) de naturaleza lingüística. De ahí que el cambio de aspecto sea el resultado de una interpretación (que es siempre, para Wittgenstein, un acto de naturaleza lingüística). Ahora bien, como hemos dicho, la interpretación se produce mediante la comparación entre figuras hasta detectar parecidos objetivos. Por ello, el ver como es un «ver la figura» más la percepción de su parecido con otras cosas. Esquemáticamente:

ver como = ver la figura + ver parecidos con otras cosas.

Por ello, el que en el ver como detectemos un elemento ligado a nuestro aprendizaje del lenguaje (un elemento conceptual) no implica que el ver como sea de naturaleza subjetiva, o que introduzca elementos extras que no están en el mundo (del tipo de los «gestalten» de Köhler), y ello es así porque lo que aprendemos a través del lenguaje es a detectar (y explotar) parecidos objetivos entre cosas.

REFERENCIAS

Obras de Witgenstein:
TLM: *Tractatus Logico Metaphisicus*.
VWVC: Gordon Baker (ed) 2003, *The Voices of Wittgenstein*, The Viena Circle, Routledge.
BPP: *Bemerkungen über die Philosophie der Psychologie*, Remarks on the Philosophy of Psychology, Basil Blackwell, 1980 (Observaciones sobre Filosofía de la Psicología, trad. de Luís Felipe Segura, UNAM, 1997)
WLPP: *Wittgenstein´s Lectures on Philosophical Psychology*, 1946-1947. Notes by P.T. Geach, K.J. Shah and A.C. Jackson 1988, Pearson Ed. Ltd. (Lecciones de Filosofía de la Psicología 1946-1947, trad. de Isidoro Reguera y Andoni Alonso, A. E. 2004)
PU: *Philosophische Untersuchungen* 1953, («Investigaciones Filosóficas», trad. De García Suárez y Ulises Moulines, Crítica, 1988)
LWPP: *Last Writings on the Philosohy of Psychology*, Basil Blackwell, 1982.

Otros:
Budd M. 1987, *Wittgenstein on Seeing Aspects*, Mind 96, 1-17.
Cavell S. 1979, *The Claim of Reason: Wittgenstein, Skepticism, Morality and Tragedy*, Oxford U.P. Oxford.
Day W. y Victor J. Krebs (eds.) (en fase de publicación): *Seeing Wittgenstein Anew*, Cambridge U.P.
Mulhall S. 1990, *On Being in the World: Wittgenstein and Heidegger on Seeing Aspects,* Londres, Routledge
Mulhall S., 'Seeing Aspects', in H. J. Glock (ed.), *Wittgenstein: A Critical Reader* (Oxford: Blackwell, 2001)
Peacocke C. 1983, *Sense and Content*, Oxford.
Schroeder S. (20), 'A Tale of Two Problems: Wittgenstein's Discussion of Aspect Perception' en Cottingham y Hacker (eds.) *Mind, Method and Morality: Essays in Honour of Anthony Kenny*, Oxford, Oxford U.P., 352-371

PHILOSOPHY AND METHODOLOGY OF SCIENCE

FILOSOFÍA Y METODOLOGÍA DE LA CIENCIA

NATURAL KIND THINKING IN PHILOSOPHY OF SCIENCE: A META-THEORETICAL ANALYSIS

ALBA AMILBURU
Department of Logic and Philosophy of Science, EHU/UPV

ABSTRACT: In a recent paper, Thomas Reydon (2010) argued for the importance of natural kind theory in philosophy of science and presented a meta-theoretical analysis of the current debate on natural kinds. Reydon identifies two lines of work or traditions that interpret and address the same philosophical problem differently, as each line of work is accompanied by a different set of assumptions and ideas. This paper is divided into two parts. In the first part, I shall present and examine critically Reydon's proposal and I shall point out certain aspects of his analysis that, should be improved. In particular, I consider two main problems on Reydon's analysis that show how this proposal is insufficient to account for the difficulties and disputes arising in the current debate on natural kinds. In the second part, I shall present an alternative proposal of a meta-theoretical analysis, which includes and develops some aspects of Reydon's analysis but introduces new elements in order to overcome its limitations. This new meta-theoretical analysis is based on a distinction between a metaphysical and a methodological approach, and I want to argue that this proposal is best suited a) for understanding the peculiarities and differences among theoretical approaches in the current philosophical discussion on natural kinds, i.e. the actual debate, and b) for explaining in what sense the notion of "natural kind" is ambiguous.
KEYWORDS: natural kind, epistemology, metaphysics, analysis

In a recent paper, Thomas Reydon (2010) argued for the importance of natural kind theory in philosophy of science and presented a meta-theoretical analysis of the present-day debate on natural kinds. Reydon identifies two lines of work or traditions that interpret and address the same philosophical problem differently, as each line of work is accompanied by a different set of assumptions and ideas.

One of them is what Reydon calls a "metaphysical approach", where natural kinds are conceived as real kinds that *exist* in nature independently of observation and human reasoning. In this classical tradition, the central problem, according to Reydon, is a metaphysical one: considering that kinds exist (in a sense) "out there" and assuming that each natural kind is defined by its corresponding real essence, the theoreticians of the metaphysical approach (see, for instance, Ellis 2001 or Garvey 2007) face issues such as how we should conceive and develop the idea of natural kindhood and real essence, or which sorts of real essence correspond to the different types of natural kind that exist.

The other tradition or line of work identified by Reydon is more recent (originated in

Bristish empiricism), and it concerns what he calls an "epistemological approach". In this case, natural kinds are considered as groups of particulars that are made by us with the purpose of being useful in a certain context. The problem now is epistemological because the main goal is to understand what it is that makes certain groupings of things (kinds) suitable for featuring successfully in explanation and prediction. Reydon considers that authors like Richard Boyd (1999) are close to this line of work.

Reydon's proposal is of great interest because the concept of natural kind is ambiguous: a fundamental disagreement concerns how philosophers understands the idea of naturalness. Actually, to determine the *ontological status* of natural kinds has proved an issue that causes confusion and disagreement among the authors. Thus, elucidating the "implicit horizon" that accompanies the notion of natural kind inserted into each of these lines of work that conform the current debate may enable us to understand why so many controversies and disagreements arise. Hence, a meta-theoretical analysis, as proposed by Reydon, is a necessary step a) to understand the debate, b) to clarify the use and meaning of the term "natural kind" and, c) to address more effectively the question concerning the *ontological status* of these classes. Before facing questions like what natural kinds are, it is convenient to address issues such as what is meant by the term “natural kind”. The paper is divided into two parts. In the first section, I shall present and examine critically Reydon's meta-theoretical analysis and I shall point out certain aspects of his proposal that, should be improved. In particular, I consider two main problematic aspects on Reydon's analysis that show how this proposal is insufficient to account for some of the difficulties and disputes arising in the current debate on natural kinds. Firstly, Reydon's proposal does not provide an adequate characterisation of the second line of work identified in the discussion. For instance, it lacks its implicit assumptions, such as the rejection of any metaphysical consideration for addressing the *ontological status* of these classes. In addition, the main difference between both lines of work, according to Reydon, is which sort of questions is given priority, metaphysical or epistemological. However, I shall argue that the main difference (and what generates tension between both lines of work) is not what sort of questions is prior, but which conception of natural kinds is arising in each line of work (which idea of “naturalness” is underlying to each conception).

And secondly, Reydon does not formulate sufficiently the main problem for both lines of work and the connection between them. The author said that the two kind of worries or central questions in both lines are intimately connected because both lines of thinking constitute different manifestations of a unique problem. As Reydon said: “[...] it might appear that there are two separate lines of thinking about *two distinct problems.* [...] But this would, I contend, be a misunderstanding of the actual situation, as the two lines of thinking constitute different ways of

approaching the same problem (2010, 251). Thus, what explains that both lines of work are trying to solve the same problem, according to Reydon, is the inseparable connection between their central questions.

> "If the basic assumption of the metaphysical line of work – that things naturally come in kinds and that kinds exist in some sense in the world – were correct, any sufficiently elaborated metaphysical account of natural kinds would also immediately answer these epistemological questions: those kinds that can successfully feature in human reasoning are precisely those that exist in nature. Conversely, any acceptable theory that answers these epistemological questions must explicate in which respects the kinds that feature in human reasoning represent features of the world. A reasonable assumption, after all, is that a grouping of things can only successfully feature in human epistemic practices if it is (at least in a minimal sense) grounded in the state of affairs in the world "out there""(2010: 251).

However, this explanation is not sufficient and adequate for understanding the "same problem" and the connection that Reydon identified in the debate. In the second section, I shall present an alternative proposal of a meta-theoretical analysis, which includes and develops some aspects of Reydon's analysis but introduces new elements in order to overcome its limitations. This meta-theoretical analysis is also based on a distinction between a metaphysical and a epistemological-methodological approach, and I want to argue that this is best suited to the peculiarities and differences among theoretical approaches in the current philosophical discussion on natural kinds. In contrast to Reydon's proposal, this analysis provides conceptual resources first, to explain in what sense the notion of "natural kind" is ambiguous, and second, to characterise both lines of work, and the relationship between them. Moreover, I show that both lines of work, in terms of this analysis are a constant in the debate (as this tension can also be seen in the works of Mill 1843, Peirce 1903, and Hacking 1991, Boyd 1991). Understanding the philosophical discussion of natural kinds in terms of this meta-theoretical proposal facilitates the interpretation and characterisation of many of the most controversial aspects of current debate.

REFERENCES

Boyd, R. 1991, 'Realism, anti-foundationalism and the enthusiasm for natural kinds'. *Philosophical Studies* 61, 127–148. Boyd, R. 1999, 'Kinds, complexity and multiple realization'. *Philosophical Studies,* 95, 67–98. Ellis, B. 2001, 'Scientific *essentialism'.* Cambridge University Press, Cambridge.

Garvey, B. 2007, 'Philosophy *of biology'.* Acumen, Stocksfield. Hacking I. 1991, 'A tradition of natural kinds'. *Philosophical Studies* 61, 109–126.

Mill, J. S. 1843, 'A *System of Logic,* in *Collected Works of John Stuart Mill',* J. M. Robson, de., Toronto: University of Toronto Press. Peirce, C. S. 1903, 'Kinds', in *Baldwin's Dictionary of Philosophy and Psychology,* New York: Ginn.

Reydon, T., 2010, 'Natural Kind Theory as a Tool for Philosophers of Science'. Suarez, M. *et al* (Eds.) *EPSA Epistemology and Methodology of Science: Launch of the European Philosophy of Science Association.* 245- 254.

EXTENSIÓN DEL DOMINIO DE LA TEORÍA DE LA SELECCIÓN NATURAL: ENFOQUES SINCRÓNICOS Y DIACRÓNICOS DEL CAMBIO CULTURAL

LORENZO BARAVALLE
FFLCH, Universidade de São Paulo
lorenzo_baravalle@yahoo.it

ABSTRACT: The aim of this paper is to present some attempts to expand the (neo-)Darwinist thesis to cultural contexts, under the categories of *synchronic approaches* and *diachronic approaches*. On the one hand, theories over the first kind of approach limit their research to the psychological traits that reinforce or hinder the spread of adaptive behaviors in humans. Here, the way researchers have been able to formulate reliable scientific models is going to be shown. On the other hand, studies concerning to the diachronic approaches try to define the processes of cultural change as such, in selective terms. We are going to argue that the assumed heterogeneity of the proposals should not discourage the search for *Guide Principles* for the development of a unitary theory.
KEYWORDS: Natural selection, Synchronic approaches, Diachronic approaches, Guide Principles, Historical explanation.

1. Enfoques sincrónicos y diacrónicos

El punto de partida de la discusión, compartido por los diferentes modelos, es el problema implicado por las conductas consideradas *maladaptativas*, a saber, conductas, suficientemente comunes y constantes en una sociedad, que implican una disminución de la *fitness* de los individuos que la componen. Una posible solución para este problema es la de pensar la difusión de estas conductas como una consecuencia indirecta, y en gran medida inevitable, de nuestras facultades psicológicas adaptadas. En otras palabras, nuestras mentes, originariamente adaptadas para considerar apetecibles determinadas conductas positivamente correlacionadas con el incremento de *fitness*, tenderían a ser atraídas, en sociedades más avanzadas respecto a las primitivas, por conductas no adaptativas o, incluso, dañinas para el individuo (Cavalli-Sforza & Feldman 1981). Esta explicación es plausible, pero insuficiente, ya que, como ha notado justamente Sober, "el modelo describe las *consecuencias* del atractivo de una idea, no la *causa* que la hizo atractiva" (1996, 339).

Sin rechazar la propuesta de Cavalli-Sforza y Feldman, es posible profundizar su análisis en dos direcciones, que corresponden a dos interpretaciones distintas del término "causa" en la cita de Sober. Si consideramos las causas de la difusión de conductas no adaptativas, o maladaptativas, como *causas próximas* (Mayr, 1961), entonces procederemos a un estudio *sincrónico* de los procesos cognitivos que, en cada momento, han determinado la preferencia hacia una determinada

conducta. La finalidad de este enfoque es la de determinar las condiciones por las cuales existe variabilidad cultural, dadas unas constantes biológicas. Por otro lado, si las causas de la difusión de una conducta no adaptativa o maladaptativa son consideradas como *causas remotas*, es plausible un análisis *diacrónico*, es decir, un análisis que determine si existen presiones selectivas que actúan históricamente sobre las conductas sociales, no adaptativas o maladaptativas, independientemente del substrato biológico.

Es significativo que, mientras los análisis sincrónicos parecen apoyarse en varias evidencias empíricas y en datos provenientes de varias disciplinas científicas (ejemplificados por los trabajos de los psicólogos evolutivos – en particular, Tooby & Cosmides, 1992), el enfoque diacrónico ofrece a menudos solo esbozos de teorías (con las excepciones de Boyd & Richerson 1985; 1992 y Richerson & Boyd 2005, aquí considerados como modelos estándar para ese tipo de enfoque, junto a Harris 1979).[1] La razón de tal desequilibrio se ha de buscar en los diferentes papeles que juega la teoría de la selección natural en los dos tipos de enfoques. Los análisis sincrónicos se fundamentan en un *principio guía*, o regulativo (o *meta-hipótesis*; Resnik 1997, 42), que es propio de los enfoques evolucionistas biológicos tradicionales, y que podría ser enunciado como:

> Principio Guía de la Selección Natural (PGSN): "cuando un rasgo muda en una población, considera su fuerza adaptativa, a saber, alguna función o conducta que, en un determinado ambiente, incrementa el éxito reproductivo (o sea, la *fitness*) mejor que si el rasgo fuera ausente" [2] (Díez & Lorenzano, *mecanuscript*, 12).

Las conductas maladaptativas son consideradas como *free riders* que no invalidan el principio, ya que la ausencia de las facultades psicológicas que las vuelven posibles sería peor, en términos de *fitness*, que la presencia de la conducta maladaptativa.

1

 Los trabajos de Richerson & Boyd, así como los de Cavalli-Sforza & Feldman (1981) y Cavalli-Sforza (2000) muestran, en realidad, una duplicidad de enfoque, con respecto a lo que aquí llamamos de "sincrónico" y "diacrónico". En otras palabras, estos autores aceptan (y engloban en sus respectivos modelos) muchos resultados de la psicología evolutiva, pero no consideran esta teoría lo suficientemente completa y la amplían, por tanto, con consideraciones genuinamente "diacrónicas".

2

 Nótese que la aplicabilidad de este principio guía a las explicaciones sincrónicas de las conductas sociales no implica que éste tenga relevancia exclusivamente en tal contexto. En ámbito estrictamente biológico puede fundamentar explicaciones tanto sincrónicas (o sea, del por qué un rasgo es *adaptativo* en la actualidad; Sober 1996, 146-7) como diacrónicas (a saber, del por qué un rasgo es una *adaptación*, por causa de un proceso que tuvo lugar en el pasado) y, de hecho, estas últimas son mucho más interesantes, desde el punto de vista de un análisis evolutivo.

En cambio, el principio guía al que pueden apelar los análisis diacrónicos es mucho más débil y no es estrictamente darwinista, ya que no es evidente que las supuestas presiones selectivas que actúan históricamente, sobre las conductas sociales, produzcan un incremento de *fitness*. Podría ser enunciado de esta manera:

> Principio Guía de la Selección Cultural (PGSC) "Cuando una conducta muda en una población, considera su fuerza adaptativa, a saber, alguna función que, en un determinado ambiente social, incrementa *x* mejor que si la conducta fuera ausente".

La *x* se refiere, según el contexto, a rasgos fenotípicos sociales (es decir, compartidos dentro de una población) complejos, que evolucionan cumpliendo funciones no directamente compatibles con el incremento de *fitness* biológica. El problema con esta formulación es que puede parecer que los enfoques diacrónicos se fundamentan en una mera tautología. Sin embargo, tentaré mostrar que la "vacuidad" del principio guía, la elección aparentemente arbitraria de los parámetros adaptativos y la subdeterminación (provisoria) del contenido empírico pueden no ser razones determinantes para rechazar este tipo de enfoques.

2. Evolución y explicación histórica

Una razón por la que los enfoques diacrónicos parecen ser vulnerables a las críticas que acabamos de exponer es identificable, en buena medida, con el difuso prejuicio hacia el tipo de explicación histórica que estarían supuestamente proponiendo. Este prejuicio, debido a la comparación de cierto pensamiento evolucionista con la creencia en "leyes históricas" (Popper 1957, § 27) o, al contrario, a su reducción a mera crónica de hechos (Fodor & Piattelli Palmerini 2010, cap. 7), ha sido ocasionalmente ampliada a la teoría de la selección natural *in toto* (recientemente, por Fodor mismo). Mi estrategia será aquí entonces, primero, la de aclarar por qué estas críticas son irrelevantes en relación a la explicación evolutiva en su dominio aplicativo tradicional – el biológico –, para ampliar, después, estas consideraciones a los modelos seleccionistas diacrónicos de la cultura. No es mi intención, por obvias razones de espacio, desarrollar un análisis sobre el carácter nomológico de la biología evolutiva, sino solo evidenciar el hecho de que la existencia de principios regulativos no empíricamente falsables (como PGSN) no perjudica el valor explicativo de las hipótesis elaboradas en su marco, como creen sus críticos, sino todo lo contrario. Y eso vale también para PGSC y para las extensiones de la teoría de la selección natural a los ámbitos culturales, aunque en un sentido ligeramente diferente.

La idea de que una explicación diacrónica, en el sentido discutido en la sección anterior – a saber, concerniente a las causas remotas –, implique la referencia a algún tipo de ley histórica y la

idea, aparentemente opuesta, de que no sea nada más que una explicación contingente son, en realidad, profundamente interrelacionadas. Ambas se basan en la premisa según la cual las leyes supuestamente invocadas para explicar, subsumiendo, los fenómenos en cuestión serían "leyes naturales de sucesión" (Popper 1957, 97), las cuales, por lo que sabemos, no existen y, por lo tanto, no pueden subsumir absolutamente nada. Según esta perspectiva, en el caso de la teoría de la evolución, las explicaciones adaptacionistas harían referencia, en el peor de los casos, a una "ley de la evolución de las formas orgánicas" (Huxley, 1880, 214 *apud* Popper 1957, 99-100) o, en el mejor, a regularidades tan generales que resultarían triviales, por lo que no deberíamos considerarlas leyes (Fodor & Piattelli Palmerini 2010, 132-5). Dada esta situación, cuando enumeramos las causas de un evento histórico, no estaríamos autorizados a decir que se trata de una explicación, sino simplemente de una narración.

Es obvio que, si la situación fuera realmente esta, no sólo las conductas sociales, sino también los rasgos biológicos serían ampliamente indeterminados con respecto a nuestra historia evolutiva. La única manera para explicar nomológicamente el funcionamiento de nuestro organismo sería a través de leyes de la física, de la química o de la biología molecular. Afortunadamente, ese punto de vista sobre el potencial explicativo de la selección natural parece ser fruto de una doble confusión. Por un lado, como explica Sober: "Para decidir si algo es una ley o una hipótesis histórica [o, añado yo, un principio regulativo] debe ser claro qué *proposición* se desea clasificar" (1996, 45). Es decir, en una teoría que quiere describir series de eventos, no necesariamente todas las proposiciones se referirán a procesos históricos. Al contrario, normalmente sólo lo harán las hipótesis concretas, las cuales, obviamente no son leyes. En este sentido, la biología se encuentra en la misma situación de la física. En suma, para explicar un evento histórico, no hace falta una ley histórica. Las leyes que encontramos en la biología evolutiva (como la de la proporción de los sexos de Fisher o la de Hardy-Weinberg), y que describen adecuadamente ciertos procesos históricos, no se refieren directamente a los procesos mismos, sino a regularidades matemáticas a las que estos procesos obedecen. [3]

Por otro lado, los que hemos definido como principios guía o regulativos, o meta-hipótesis, no son, como han insistido varios autores, con matices diferentes (Sober 1996; Resnik 1997; Diez & Lorenzano *mecanuscript*), propiamente leyes (las cuales permitirían la formulación de hipótesis falsables), sino reglas *heurísticas* que estimulan la elaboración de modelos en un determinado

3

 Esta es, al menos, la posición de Sober, aunque la literatura abunda de otras defensas del concepto de ley en biología, en diferentes sentidos y no siempre compatibles, que aquí no cabe discutir (por ejemplo Brandon 1997; Waters 1998; Elgin 2006).

contexto, a saber, el del adaptacionismo. El error de los que critican las explicaciones evolutivas por su carácter histórico sería, en este sentido, el de confundir estas reglas con hipótesis concretas sobre procesos históricos y, por lo tanto, considerarlas como metafísicas o vacías. En realidad, se trataría de generalizaciones que funcionan como *esquemas* de leyes (como también es posible encontrar en la física; Kuhn 1970 *apud* Diez & Lorenzano *mecanuscript*) que direccionan la formulación de hipótesis contrastables, sin ser ellas mismas contrastables.

3. ¿Qué aportan los enfoques evolutivos diacrónicos?

Una vez que quede claro que no son los principios guía los que se prestan a falsación, sino los modelos elaborados en su marco, volvamos a considerar la diferencia entre PGSN, aplicado tanto a la selección natural en el ámbito tradicional como a la selección sincrónica de rasgos culturales, y PGSC. No podemos decir que el adaptacionismo implicado por PGSN, a saber, la tendencia al incremento de *fitness*, goza de mayor apoyo empírico que la versión modificada presentada en PGSC, ya que, en principio, estas meta-hipótesis no se ofrecen a contrastación. En cierto sentido, PGSC *no es más vacuo que* PGSN. Sin embargo, se podría decir que los modelos presentados bajo el marco de PGSN son apoyados por un esquema de leyes, mientras los modelos propios de PGSC no, ya que en PGSC no está especificado ningún parámetro propio de adaptación. Eso equivaldría a decir que la *indeterminación de los parámetros adaptativos* en PGSC compromete irremediablemente cualquier caracterización nomológica, aunque débil, de los modelos diacrónicos.

No obstante, existe un sentido en el que los modelos evolutivos diacrónicos del cambio cultural expresan regularidades. En ellos no se cumplen necesariamente las condiciones que correspondan a la optimización de la *fitness*, sin embargo la difusión y la fijación de los elementos mínimos de la transmisión cultural (las *variables culturales*, en la terminología de Richerson & Boyd 2005) son analizables por medio de conceptos evolutivos clásicos: variabilidad (mutación), hereditariedad y deriva, entre otros (Abrantes & Portela L. Almeida, 2011). Algunos estudios que ejemplifican (no siempre explícitamente) ese tipo de enfoque son: el de las prohibiciones alimentares en la India y en el Islam de Harris (1998), de la baja natalidad en determinadas clases de las sociedades occidentales contemporáneas, de Cavalli-Sforza y Feldman (1981), y Richerson y Boyd (2005), de la dinámica y la resolución del conflicto entre Nuer y Dinka (originariamente estudiado por Kelly 1985, y tratado en términos evolutivos por Sober & Wilson 1998 y Richerson & Boyd 2005). Sober y Wilson (1998) han sugerido que estas regularidades evolutivas son el resultado de selección de grupo, o de selección en múltiples niveles. Aunque concuerde en general con el análisis de estos autores (por los que la diferencia entre PGSN y PGSC estaría simplemente en el

énfasis atribuido al incremento de la *fitness* individual o del grupo), sostengo que, al menos idealmente, se podría ir más allá: la selección de conductas culturales complejas conduce a adaptaciones que no son únicamente definibles en términos de incremento de la reproducción.

Esta afirmación, que resultaría aparentemente incontrastable desde un punto de vista empírico, ya que, como hemos vistos, no especifica un parámetro medible, como la *fitness*, puede en realidad adquirir una forma científicamente aceptable en los modelos particulares. Harris, por ejemplo, propone parámetros adaptativos fundados sobre la optimización de la productividad de las poblaciones; Richerson y Boyd parecen indicar la posibilidad de procesos de maximización de la cohesión social, en un sentido no progresivista. Es obvio que la plausibilidad de estos modelos encuentra límites importantes en la dificultad de determinar cuáles son los rasgos conductuales *bona fide* que serían culturalmente seleccionados, a partir de los criterios definidos en su interior. Y es posible que la crítica por la cual estos modelos son aun ampliamente *subdeterminados*, desde el punto de vista empírico, sea justificada. Sin embargo, esto no tiene por qué implicar un rechazo *a priori* de los modelos diacrónicos: el hecho de que los grupos humanos evolucionen a través de presiones selectivas independientes de la optimización de la *fitness* debería ser investigado empíricamente.

En este sentido, y para concluir la discusión, la diferencia entre enfoques sincrónicos y enfoques diacrónicos del cambio cultural debería ser considerada como la diferencia entre dos programas de investigación. Los primeros se colocan por completo dentro de una tradición darwiniana, la cual, como hemos tenido la ocasión de reafirmar en esta presentación, no tiene nada que ver con una teoría sobre "leyes de la evolución" ni es reducible a la historia natural. Los segundos se colocan fuera de lo que, hasta ahora, ha sido considerado como el darwinismo estándar. Podría surgir la duda de que quieran revivir tendencias historicistas, pero en realidad sólo avanzan hipótesis sobre la aplicabilidad de los conceptos evolutivos al desarrollo de las sociedades humanas y a sus productos. En este sentido, simplemente favorecen la introducción de herramientas darwinianas en el ámbito de las ciencias sociales. Aunque PGSC no ofrezca un marco unitario para estos enfoques, esto no significa que en un futuro no podrá hacerlo.

Agradecimientos

Este trabajo ha sido realizado con la ayuda de la Fundação de Amparo à Pesquisa do Estado de São Paulo – Brasil –, a la que agradezco el apoyo y la financiación de mi proyecto de pesquisa "Estratégias cognitivas e evolução humana" (Processo 10/17862-7).

REFERENCIAS

Abrantes, P.C. & Portela L. Almeida, F. 2011, 'Evolução humana: a teoria da dupla hernça', en Abrantes, P.C. et al. (ed.), *Filosofia da Biologia*, Porto Alegre, Artmed.
Boyd, R. & Richerson, P. 1985, *Culture and the evolutionary process*, Chicago, University of Chicago Press.
Boyd, R. & Richerson, P. 1992, 'Punishment Allows the Evolution of Cooperation (or Anything Else) in Sizable Groups', *Ethology and Sociobiology*, 13, 171-95.
Brandon, R.N. 1997, 'Does Biology Have Laws? The Experimental Evidence', *Philosophy of Science*, 64, 444-57.
Cavalli-Sforza, L. 2000, *Genes, pueblos y lenguas*, Barcelona, Crítica.
Cavalli-Sforza, L. & Feldman, M. 1981, *Cultural transmision and evolution: a quantitative approach*, Princeton, Princeton University Press.
Diez, J.A. & Lorenzano, P. Who got what wrong? Fodor & Piattelli and Sober on Darwin: Guide Principles and explanatory models in natural selection. *Mecanuscript*.
Elgin, M. 2006, 'There May Be Strict Empirical Laws in Biology, after All', *Biology and Philosophy*, 21, 119-34.
Harris, M. 1979, *Cultural Materialism: The Struggle for a Science of Culture*, New York, Random House.
Harris, M. 1998, *Good to Eat: Riddles of Food and Culture*, Illinois, Waveland Press.
Mayr, E. 1961, 'Cause and effect in biology', *Science*, 134, 1501-6.
Huxley, T.H. 1880, *Lay Sermons*.
Kelly, R. 1985, *The Nuer Conquest: The Structure and Development of an Expansionist System*, Ann Arbor, University of Michigan Press.
Kuhn, T.S. 1970, 'Second Thoughts on Paradigms', en Suppe, F. (ed.) 1974, *The Structure of Scientific Theories*, Urbana, University of Illinois Press, 459-82.
Popper, K. 1957, *The Poverty of Historicism*, London, Routledge and Kegan Paul.
Resnik, D. 1997, 'Adaptationism: Hypothesis or Heuristic', *Biology and Philosophy*, 12, 39-50.
Richerson, P. & Boyd, R., 2005, *Not by Genes Alone: How Culture Transformed Human Evolution*, Chicago, Chicago University Press.
Sober, E. 1996, *Filosofia de la biología*, Madrid, Alianza editorial, 1996.
Sober, E. & Wilson, D.S. 1998, *Unto Others. The Evolution and Psychology of Unselfish Behavior*, Cambridge, Massachusetts, Harvard University Press.
Tooby, J. & Cosmides, L. 1992, 'The Psychological Foundations of Culture' en Barkow, J.H., Tooby, J. & Cosmides, L. (ed.), *The Adapted Mind: Evoltionary Psychology and the Generation of Culture*, New York, Oxford University Press, 19-139.
Waters, C.K. 1998, 'Causal Regularities in the Biological World of Contingent Distributions', *Biology and Philosophy*, 13, 5-36.

LAS ANOMALÍAS NO REFUTATORIAS Y EL ALCANCE DEL DOMINIO TEÓRICO*

MARÍA CAAMAÑO ALEGRE
Universidad de Valladolid
mariac@fyl.uva.es

ABSTRACT: The present paper is focused on Laudan's notion of non-refuting anomaly and its relevance for the question of empirical adequacy, ultimately connected to that of a theory's domain. Non-refuting anomalies, in contrast to refuting ones, do not involve any logical incompatibility between empirical consequences of the theory on the one hand and verified statements regarding empirical facts on the other. They rather entail a theory's incapability to account for certain kind of salient empirical phenomena whose description is consistent with everything established by the theory. Non-refuting anomalies, therefore, do not primarily point to any mistake on the way in which a theory explains the phenomena but rather to the incompleteness on the part of the theory. The paper is structured in three parts: first, it is argued that the acknowledgment of non-refuting anomalies leads to the rejection of a consequentialist theory of evidence *à la* van Fraassen; second, the notion of the applicative domain of a theory is reconsidered so that the "intentionalist" model is replaced by one where salience plays a crucial role; and third, the significance of non-refuting anomalies is illustrated by two historical cases. The following two examples will be discussed: Laplace's introduction of the nebular hypothesis in order to account for the fact that all planets move in the same direction and nearly in the same plane, and non-refuting anomalies involved in the geological revolution towards the middle of the twentieth century.
KEY WORDS: non-refuting anomaly, empirical adequacy, applicative domain

El presente trabajo se centra en las implicaciones que la noción de anomalía no refutatoria tiene para el análisis de la adecuación empírica y, en última instancia, para el esclarecimiento de la noción de dominio teórico. Se argumentará que el reconocimiento del éxito teórico asociado a la resolución de anomalías no refutatorias, caracterizable como una mejora en el alcance de las teorías, sólo puede explicarse tras haber reconsiderado previamente la naturaleza del dominio de aplicación de éstas. Por su propio carácter, este tipo de anomalías no resultan identificables a partir de las predicciones fallidas o consecuencias empíricas falsas de la teoría (opción consecuencialista), ni atendiendo a las aplicaciones pretendidas de ésta (opción pragmático-intencionalista). Se requiere, por tanto, una concepción más amplia del dominio de aplicación, donde tenga cabida y posea sentido la noción de anomalía no refutatoria. En este trabajo se tratará de esclarecer la noción de dominio teórico a partir de la elucidación de la propia noción de anomalía no refutatoria. Para ello, en primer lugar, se explicará la relevancia de las anomalías no refutatorias en relación con la

* Este trabajo ha sido financiado con cargo a los siguientes proyectos de investigación: *Lenguaje y pensamiento: dependencia contextual*, FFI2008-06421-CO2-02/FISO, y *Actitudes cognitivas y la justificación del conocimiento*, FFI2009-08828/FISO, ambos concedidos por el Ministerio de Ciencia e Innovación.

cuestión de la adecuación empírica, mostrándose las limitaciones de la noción consecuencialista de adecuación empírica manejada por van Fraassen (1992, 2008, 247-254). En segundo término, se argumentará que el reconocimiento de las anomalías no refutatorias obliga a una caracterización más amplia del dominio de aplicación de las teorías, donde se tomen en consideración, como delimitadores de dicho dominio, no sólo los aspectos consecuencialistas y pragmático-intencionales, sino también aquellos concernientes a la prominencia (*salience*) de las regularidades empíricas caracterizables mediante los conceptos empíricos de las teorías (Nickles, 1988). En la tercera y última parte, se examinan un par de casos históricos en los que queda patente, tanto el papel jugado por las anomalías no refutatorias, como sus implicaciones por lo que respecta al dominio aplicativo de las teorías. Se analizan dos casos distintos: uno en el que la teoría enfrentada a tales anomalías se completa de forma que llega a resolverlas, como ocurre con la hipótesis nebular de Laplace (Brush, 1996, Laplace, 1796/1930) y otro en el que las teorías enfrentadas a las anomalías se rechazan y reemplazan por otras capaces de resolver las anomalías, como sucedió, en geología, con el paso de la teoría contraccionista a la de la deriva continental, y de esta a la teoría de las placas tectónicas (Laudan, 1977, 29 y 2000, 167, Cohen, 1985/2001).

1. Las anomalías no refutatorias y su importancia para la cuestión de la adecuación empírica

Haciendo leves modificaciones a la concepción laudaniana,[1] las anomalías no refutatorias se entenderán aquí como aquellos fenómenos empíricos prominentes, relevantes para la teoría, cuya descripción es consistente con lo establecido por ésta, pero cuya explicación no resulta posible desde la teoría en cuestión. En su obra de 1977, Laudan menciona algunos ejemplos, como la incapacidad de la cinemática pregalileana para explicar los rasgos matemáticos del movimiento pendular y la falta de explicación, desde la mecánica newtoniana, del hecho de que todos los planetas orbiten en la misma dirección y aproximadamente en el mismo plano. En su texto de 2000, la lista de casos ilustrativos se amplía considerablemente:

> "For instance, stable-continent theories of geology offered no explanation as to why the continents fit together so neatly. (…) Steady state cosmology offered no explanation for residual background radiation. (…) Ptolemaic astronomy did not explain why – even within Ptolemy's own theory – all the planets have a solar component to their motion. Phlogistic chemistry is wholly silent about why gaseous elements combine only in integral multiples by volume. Theories of terrestrial causes of dinosaur extinction leave unexplained the worldwide iridium spike that

1 Recordemos que Laudan caracteriza las anomalías en general como aquellos problemas empíricos que suscitan dudas razonables acerca de la adecuación empírica de una teoría una vez que otra teoría los ha solucionado (cf. 28, 30). La peculiaridad de las anomalías no refutatorias radicaría en que pondrían de manifiesto la incapacidad de una teoría para dar cuenta de unos fenómenos empíricos cuya descripción es consistente con todo lo establecido por la teoría (Laudan, 1981, p. 618).

> occurred towards the end of the Cretaceous. Geostatic models of the Earth cannot explain the Coriolis effects associated with large bodies of wind and water" (p. 167).

Debe aclararse que, en el presente aproximación al tema, se prescindirá de un aspecto que Laudan incluye en la caracterización de las anomalías no refutatorias, a saber, la idea de que éstas sólo pueden ser consideradas como tales una vez que otra teoría ha podido resolverlas. Se evitará en lo que sigue mantener esta restricción en relación con las anomalías no refutatorias, favoreciendo una caracterización más amplia de acuerdo con la cual éstas constituyan problemas empíricos que conduzcan a cuestionar la adecuación empírica de la teoría, con independencia de si alguna otra teoría ha logrado resolverlos. Con todo, la aparente desconexión lógica entre este tipo de anomalías y las teorías suscita el interrogante acerca de la plausibilidad de considerar a las primeras como formando parte del dominio de aplicación de las segundas. Se plantea, por tanto, la cuestión relativa al criterio conforme al cual delimitar el dominio aplicativo de las teorías. La justificación empírica de una teoría depende directamente de lo que se reconozca como formando parte de su dominio de aplicación, por lo que, en el caso de las anomalías no refutatorias, su valor para la evaluación empírica de una teoría dependerá de que se proporcionen buenas razones para entender que dichas anomalías deben reconocerse como parte del dominio aplicativo. Según la concepción de éste que se asuma, la propia noción de adecuación empírica se concebirá de distintas formas. Como advierte Laudan, si entendemos la adecuación empírica exclusivamente en el sentido de corrección observacional, entonces difícilmente podremos discriminar entre teorías con mayor o menor éxito teórico, puesto que teorías de muy distinto alcance podrían ser todas ellas observacionalmente correctas. Más aún, teniendo en cuenta la relación inversa que suele darse entre la probabilidad y el alcance de una teoría, si sólo nos interesase la verdad o la alta probabilidad con independencia del alcance de la teoría, tendríamos que conformarnos con modestas generalizaciones empíricas o afirmaciones sobre hechos particulares renunciando a la teorización (2000, 170-2).

El reconocimiento de las anomalías no refutatorias exigiría, por tanto, la superación del modelo consecuencialista de relevancia evidencial, que Laudan describe en los siguientes términos:

> "Consequentialism, whose best known variants are called inference to the best explanation and hypothetico-deduction, is the view that the plausibility or acceptability of an hypothesis or theory is determined exclusively by an examination of the truth of the accessible consequences of that hypothesis or theory" (1995, 28)

Oponiéndose a esta idea, Laudan argumenta que ser una consecuencia de una hipótesis no es una condición necesaria ni suficiente para que pueda considerarse como evidencia para dicha hipótesis. La adecuación empírica de una teoría, por tanto, no sólo dependerá de la verdad o

falsedad de sus consecuencias empíricas, sino también de su correspondencia con los fenómenos más prominentes de su dominio de aplicación. Expresado en términos modelo-teóricos: la adecuación empírica de una teoría no requiere simplemente que (en al menos uno de sus modelos) todas sus subestructuras empíricas sean isomorfas con los fenómenos correspondientes (van Fraassen, 1976), sino también que sean isomorfas con todos los fenómenos prominentes de su dominio. La posición consecuencialista podría tratar de defenderse asumiendo que, ante cualquier tipo de anomalía, una teoría siempre se acabará mostrando capaz o incapaz de solventarla, es decir, se revelará finalmente decidible. A esta idea podría responderse que, como evidencian las resoluciones de las anomalías no refutatorias, lo que se necesita es un crecimiento del contenido informativo de la teoría. Ello implica, bien la incorporación de nuevos conceptos y postulados a la teoría inicial (creando por tanto una nueva teoría), bien el rechazo de la teoría inicial y su sustitución por una teoría completamente nueva.

2. Una caracterización más amplia del dominio de aplicación de una teoría

Surge, pues, como cuestión central el problema de cómo distinguir las anomalías no refutatorias de casos en los que ciertos fenómenos empíricos permanecen inexplicados por una teoría, bien por el mero hecho de que caen fuera de su dominio de aplicación, bien porque, a pesar de hallarse incluidas en dicho dominio, se trata de fenómenos que la misma teoría podrá explicar en el futuro una vez se disponga de más datos. A diferencia de lo que ocurre en el primer caso, las anomalías no refutatorias presentan un alto grado de similitud o una estrecha conexión con los ejemplares o aplicaciones intencionales de la teoría. En contraposición a lo que sucede en el segundo caso, este tipo de anomalías, debido a su carácter recalcitrante, sugieren que la teoría necesitará modificaciones sustanciales, adiciones o incluso una sustitución. Debe subrayarse que las anomalías no refutatorias apuntan a casos en los que la teoría permanece en silencio con respecto a ciertos fenómenos, y no a casos en los que se aplica si bien sólo de forma muy aproximada. Igualmente es importante advertir que las anomalías en cuestión consisten en fenómenos prominentes o llamativos que, dada su relevancia para la teoría, ésta debería poder acomodar.

La importancia de atender a la completitud o el alcance explicativos de una teoría, frente a lo que sería su mera corrección observacional, debe conducir a un tipo de análisis del dominio teórico en el que, por una parte, la naturaleza del dominio de aplicación de una teoría se entienda como determinado a partir del aparato conceptual heredado por la teoría (ya sea éste preteórico o proveniente de teorías previas presupuestas en la realización de las medidas y experimentos). Por otra parte, tanto las relaciones postuladas teóricamente como las relaciones empíricamente

determinadas relevantes para su contrastación deberían concebirse como algo dinámico y abierto a futuros desarrollos. Esta segunda tarea requerirá el uso de herramientas representacionales, como las ofrecidas desde el *partial structures approach* (Bueno, 1999), que permitan determinar los distintos componentes teóricos y empíricos explicativos del carácter incompleto de la teoría. Aquí se atenderá a la primera tarea, que puede realizarse con ayuda de la noción técnica estructuralista de modelo potencial parcial **(**Balzer, Moulines, Sneed, 1987, 4), que, a diferencia de la noción de subestructura empírica manejada por van Fraassen, no implica el carácter observacional de las estructuras ni su "incrustabilidad" en los modelos teóricos (actuales).

En primer lugar, por tanto, se requiere un análisis minucioso de los recursos conceptuales de una teoría, en particular, de aquellos que intervienen en la representación de aquellos fenómenos empíricos que constituyan anomalías no refutatorias. Una caracterización adecuada del dominio de aplicación de una teoría debe poder dar cuenta de dos tipos posibles (y no mutuamente excluyentes) de relaciones entre las aplicaciones intencionales y el aparato teórico: una lógica y la otra conceptual. Las anomalías no refutatorias guardan una relación de incrustación en el aparato conceptual empírico de la teoría o modelos potenciales parciales, pero no así en sus leyes o modelos actuales. Aplicando una distinción estructuralista,[2] podemos diferenciar entre la clase de los modelos potenciales de una teoría (Mp), los cuales satisfacen las condiciones marco de la teoría, y clase de los modelos actuales (M), que satisfacen tanto las condiciones marco como las leyes teóricas. Las condiciones marco se caracterizan por proveer simplemente el aparato conceptual de las teorías, es decir, el conjunto de propiedades formales de sus conceptos científicos característicos. El marco conceptual de una teoría consiste simplemente en los recursos descriptivos proporcionados por ella, sin decir nada sin embargo acerca del mundo. Los axiomas o condiciones satisfechas por Mp son enunciados axiomáticos no legaliformes destinados a especificar el tipo o estructura de cada concepto (esto es, de cada dominio o relación incluida en la teoría). Esta clase de axiomas, por tanto, establece qué estructura debe tener cada uno de los conceptos fundamentales de la teoría, de manera que la tupla formada por las correspondientes entidades teórico conjuntistas puedan considerarse como candidatos posibles a modelos actuales de la teoría. Las leyes teóricas, por el contrario, dicen algo acerca del mundo, y lo hacen precisamente empleando los conceptos determinados por la condiciones marco de la teoría. Una consecuencia importante que se sigue de la distinción anterior es que los modelos potenciales pueden o no ser también modelos actuales. En el segundo caso, los modelos satisfarían sólo las condiciones marco y no las leyes teóricas. La distinción entre Mp y M, por consiguiente, no es meramente analítica, sino que más bien se refiere

2 Cf. Balzer, W., Moulines, C. U., Sneed, J., (1987), 2-6, Moulines, C. U., (1991), 237, 247.

a dos clases de estructuras genuinamente diferenciadas; una de ellas (M) incluida en la otra (Mp). Otra consecuencia importante es que la determinación de aquella parte de los Mp que no se solapa con M permite evidenciar que, incluso dejando M aparte obtendríamos una clase de modelos potenciales que se mantienen por sí mismos. Atendiendo a aquellas subclases de modelos potenciales no solapables con los actuales, resulta posible explicar el hecho de que no todo desacuerdo acerca de las leyes fundamentales implica un desacuerdo acerca de cómo un cierto dominio debe ser conceptualizado por la teoría. A diferencia de lo que se propone desde el empirismo estructural de van Fraassen, donde no se presenta ningún análogo para al anterior distinción estructuralista entre Mp y M, los modelos empíricos de una teoría (entre los que habría que incluir las anomalías no refutatorias), no se obtienen eliminando los componentes teóricos de los modelos teóricos (que equivaldrían a M), sino de Mp. El resultado de esa operación de recorte sería la clase de los modelos potenciales parciales (Mpp). Las anomalías no refutatorias, por su propia naturaleza, no llegarían a determinarse aplicando la función de recorte a M, pero sí podrían identificarse en muchos casos aplicándola a Mp.

Mpp puede identificarse con la base empírica o el dominio empírico de aplicación de T. Dado que una teoría suele orientarse hacia un subconjunto particular de Mpp, desde el enfoque estructuralista se ha introducido la noción de aplicación intencional para referirse a él (I). I posee un carácter preeminentemente pragmático, ya que incluye aquellas subestructuras empíricas que constituyen el objeto de investigación en que se centra la comunidad científica que asume la teoría en cuestión. Dado que las anomalías no refutatorias evidencian las limitaciones de una teoría en relación con su alcance y contenido informativo, tales anomalías difícilmente se incluirán en I, aunque así debiera ser dada su relevancia y prominencia. Las relaciones entre las distintas clases de modelos referidas arriba quedan reflejadas en el siguiente diagrama:

Modelos actuales exitosos

Mp Mpp

Anomalías I

no refutatorias

M

Las anomalías no refutatorias se entienden aquí como modelos potenciales parciales que, dada su prominencia y relevancia para T, deberían ser incluidos en I. Tal relevancia se explicaría por

la similitud o estrecha conexión entre dichas anomalías y los modelos parciales incluidos en I. Parece razonable esperar que una teoría mecánica cosmológica dé cuenta del hecho de que todos los planetas giren en la misma dirección y en aproximadamente el mismo plano. Igualmente, parece apropiado exigir que una teoría geológica explique el hecho de que la forma de los continentes encaje tan nítidamente.

3. Algunos ejemplos históricos

Al ilustrar la noción de anomalía no refutatoria a partir de ciertos casos hitóricos, Laudan reitera dos ejemplos especialmente reveladores (1977, 29, 2000, 167). El primer ejemplo de cómo este tipo de anomalía afecta al desarrollo de las teorías es la introduction, por parte de Laplace, de la hipótesis nebular para poder explicar el hecho de la unidireccionalidad y aproximada coplanaridad de los planetas (Laplace, 1796/1930, pp. 361-363). Este hecho constituía una anomalía no refutatoria para la mecánica newtoniana. Laplace, con todo, fue capaz de resolver la anomalía sin descartar la mecánica newtoniana, simplemente realizando un añadido informativo, a saber, la hipótesis nebular. Según esta hipótesis, el Sol era originalmente una gigantesca nube de gas rotando de manera uniforme. Debido al enfriamiento y a la acción de la gravedad, la nube se fue contrayendo y acelerando. La aceleración en la rotación causaría la separación de parte del borde de la masa en rotación, que acabaría por condensarse y formar un planeta. El proceso se iría repitiendo a lo largo del tiempo hasta llegar a darse la formación planetaria que contemplamos actualmente.

El segundo ejemplo se refiere a las anomalías no refutatorias involucradas en la revolución geológica acontecida a mediados del siglo XX (Thagard, 1992, cap. 7). Anomalías geológicas como el encaje entre las líneas costeras de los continents permanecieron inexplicadas por las teorías contraccionista y sólo pudieron resolverse una vez que dicha teoría fue rechazada y reemplazada por la teoría de la deriva continental. Ésta, a su vez, se vio debilitada por toda una serie de anomalías no refutatorias que condujeron a su sustitución por la teoría de las placas tectónicas (desarrollada entre 1962 y 1968). De acuerdo con la literatura de referencia sobre el tema (Laudan, 1978, Cohen, 1985/2001, Thagard, 1992, Marvin 2001), algunas de las observaciones más relevantes involucradas en dicho cambio teórico estuvieron relacionadas con la composición del lecho marino y con la distribución no azarosa de los terremotos. En ninguna de estas transiciones entre teorías geológicas llegó a producirse una refutación de lo establecido por las respectivas leyes de las teorías. El motor del cambio teórico fueron, principalmente, anomalías no refutatorias para

las que las teorías precedentes no disponían de recursos explicativos suficientes aunque sí de recursos descriptivos adecuados para representarlas.

REFERENCIAS

Balzer, W., Moulines, C. U., Sneed, J. D. 1987, *An Architectonic for Science. The Structuralist Program*, Dordrecht, Reidel.

Brush, S. G. 1996, *Nebulous Earth: The Origin of the Solar System an the Core of the Earth from Laplace to Jeffreys*, Cambridge, CUP.

Bueno, O. 1999, "What Is Structural Empiricism? Scientific Change in an Empiricist Setting", *Erkenntnis*, 50, 59-85.

Cohen, I. B. 1985/2001, "Continental Drift and Plate Tectonics: A Revolution in the Earth Science", *Revolution in Science*, Cambridge, MA: Harvard University Press, 29, 446-466.

Laudan, L. 1977, *Progress and its problems,* Berkeley, UCP.

- 1981, "Anomalous Anomalies", *Philosophy of Science*, vol. 48, 4, 618-619.

- 1995 "Damn the Consequences!", *Proceedings and Addresses of the APA*, 69, 27-34.

- 2000 "Is Epistemology Adequate to the Task of Rational Theory Evaluation?", en *After Popper, Kuhn and Feyerabend,* Dordrecht, Kluwer Academic Publishers, 165-175.

Moulines, C. U. 1991, *Pluralidad y recursión. Estudios epistemológicos*, Madrid, Alianza Editorial.

Thagard, P. 1992 *Conceptual Revolutions*, Princeton, Princeton University Press.

Van Fraassen, B. 2008, *Scientific Representation: Paradoxes of Perspective*, Oxford, OUP.

INDUCTION, LAWS OF NATURE AND INFERENCE TO THE BEST EXPLANATION*

EDUARDO CASTRO
University of Beira Interior
LanCog, University of Lisbon
ecastro@ubi.pt

ABSTRACT: What is the best explanation of our observed regularities, such as all observed ravens are black? In this paper, two potential competing explanations on laws of nature are discussed: the timeless necessity of David Armstrong and a new brand of necessity – Helen Beebee's time-limited necessity. It will be defended, against Beebee, that the timeless necessity explanation constitutes a better explanation of our observed regularities than does the time-limited necessity explanation.
KEY-WORDS: Armstrong, Beebee, induction, inference to the best explanation, laws of nature, necessity.

Introduction

In the light of laws of nature, the problem of induction is given by the argument:

(1) All observed *F*s are *G*s.

(2) The best explanation for (1) is that it is a law of nature that all *F*s are *G*s.

(∴) All *F*s are *G*s.

The regularity theory states that laws of nature are identified with the regularities of the world. This view implies that inductive scepticism is inevitable, because 'it is a law that *F*s are *G*s' simply means 'all *F*s are *G*s', where 'all *F*s are *G*s' is the conjunction 'all observed *F*s are *G*s and all unobserved *F*s are *G*s'. Our *explanandum* is not explained by itself, neither is it explained by the unobserved *F*s. That is, (2) is not a genuine inference to the best explanation (hereafter, IBE). On the contrary, the necessitarian view of Armstrong 1983 states that laws of nature are first-order universals (*F*, *G*) necessary connected in the actual world by a second-order universal, *N*, a state of affairs, *N* (*F*, *G*). This view implies that (2) is an instance of IBE. IBE is rational. Thus, the inductive inference (2) is rational.

Beebee 2011 takes a subtle different approach concerning (2):

* This work was supported by grant SFRH/BPD/46847/2008 and the project PTDC/FIL-FCI/109991/2009, "Hilbert's Legacy in the Philosophy of Mathematics", from Fundação para a Ciência e Tecnologia.

Explanandum:

(E) All observed *F*s have been *G*s *so far*.

Armstrong's *explanans*:

(T) *F* and *G* are timelessly (eternally) necessarily connected.

Beebee's *explanans*:

(SF) *F* and *G* have been necessarily connected *so far*.

We have two *explanans* for the same *explanandum*. Beebee argues that (T) is not a better *explanans* than is (SF). If this is right, Armstrong's solution to the problem of induction is a failed solution, because an extra step – an unjustified inductive inference – is needed to lead us from (SF) to (T).

This paper is a defence of Armstrong's solution and the following objections will be discussed:

(1) (SF) is incoherent.
(2) (T) is a better *explanans* than is (SF), because:
  (2.1) (T) is simpler than (SF);
  (2.2) (T) has predictive power; (SF) has no predictive power.

1. Beebee's reply to (1)

There are time-limited necessities and these necessities can justify (SF). Beebee argues that N_t (*F*, *G*), for example, could be such relation. N_t (*F*, *G*) is the relation that any *F* prior to *t* is necessarily *G* and future *F*s may or may not be *G*s.

2. Counter-reply

Beebee claims that (a) 'maybe the Universe will start behaving in other regular (…) ways' or that (b) '*F*s might stop being *G*s for no better reason than a particular time has passed' and chaos descend upon the Universe (Beebee 2011, 504, 513). That is, according to the time-limited necessitarian, there are two slighting different ways for the changing of the laws of nature. In both cases the connection between *F*s and *G*s can stop, but in the first case, the second-order universal, which connects *F*s and *G*s, might change, starting to connect other first universals with *F*s, for

example, *F*s can start to be connected with *H*s.

The basic idea of the first notion of law-change is, for example, that an observed emerald might change colour. In this instant, I have a green emerald in my hand and, *all else being equal*, the emerald can be blue in the next instant. This scenario motivates the introduction of the *new riddle of induction* of Goodman 1954. Let grue be the predicate: X is grue if, and only if, X is green before *t* and X is blue after *t*.

We have two inductive arguments and two incompatible conclusions. (1) All observed emeralds are green; therefore, all emeralds are green. (2) All observed emeralds are grue; therefore, all emeralds are grue. The same observational evidence implies two incompatible conclusions about emeralds observed in the future: there will be green emeralds and blue emeralds.

The two incompatible conclusions of the new riddle of induction are a problem for the timeless necessitarian. With the purpose to overcome this problem, the timeless necessitarian rejects the predicate grue. Based on our science, grue does not correspond to any universal. Grue does not describe any regularity. So it is (nomic) impossible to infer the law that all emeralds are grue, since there is no second-order universal relation between the universal *emeraldness* and the pseudo-universal *gruenness*. Nonetheless, the new riddle of induction is not a problem for the time-limited necessitarian, since the incompatible conclusions are not deduced. (1) All so far observed emeralds have been green; therefore, all emeralds have been green, *so far*. (2) All so far observed emeralds have been grue; therefore, all emeralds have been grue, *so far*. For the time-limited necessitarian the future is open.

According to the time-limited necessitarian, if after *t* all emeralds turn on blue, then the law of nature 'all emeralds are green, *so far*', changes to 'all emeralds are blue, *so far*'. The universal *emeraldness* starts a new necessary connection with the universal *blueness*.

Explanandum

Before *t*, all observed emeralds had been green and
after *t*, all so far observed emeralds have been blue

↓

Explanans

Before t, *emeraldness* and *greenness* had been necessarily connected and
after *t*, *emeraldness* and *blueness* have been necessarily connected, *so far*.

Synthetically, this schema is equivalent to the following one:

Explanandum: all so far observed emeralds have been grue.

↓

Explanans: *emeraldness* and *gruenness* have been necessarily connected, *so far*.

However, the introduction of the predicate grue shows that no new connections have started. After *t*, *emeraldness* did not start to have necessary connections with *gruenness*. We simply discover a law that accommodates all past observations. We discover that *emeraldness* and *gruenness* have always been connected so far, before and after *t*.

A way out for the time-limited necessitarian is to reject the predicate grue and to insist that, after *t*, *emeraldness* and *blueness* have been necessarily connected *so far*. After *t*, a new law of nature has started. However, this way out also implies that after *t*, *emeraldness* and *greenness* have been necessarily *disconnected*, i.e., after *t*, the necessary connection between *emeraldness* and *greenness* stopped. We were driven to the second notion of change.

Does it make sense thinking that laws of nature can stop? It seems that such thing is nomical impossible. If *F* and *G* are first-order universals connected by a second-order universal, *N*, then this connection cannot change in time, since universals simply does not change in time. $N(F, G)$ and $\neg N(F, G)$ are incompatible states of affairs. If it is possible that before *t*, *F*s are *G*s, and after *t*, *F*s are not *G*s, then the law that all *F*s are *G*s sinks in a mere regularity and it is not connected by a second-order universal, *N*, that is, it is not a law of nature at all.

4. Beebee's reply to (2.1) and (2.2)

Beebee argues that, in the light of *simplicity*, the result is a tie between (SF) and (T). Is it *predictive power* an explanatory virtue? In science, yes; in metaphysics, no. In science a hypothesis that generates novel predictions is to be preferred to another that only gives explanations about past phenomena. This is a line of reasoning current in science. But our business here is metaphysics. Predictive power is not an explanatory virtue in metaphysics, neither in practical nor in theoretical sense. Moreover, the timeless necessitarian begs the question against the sceptic, since the sceptic simply denies the assumption that predictive power is an explanatory virtue. An argument for predictive power must be advanced against the sceptic.

5. Counter-reply

Simplicity

(T) is *syntactical* simpler than (SF), since (SF) depends of time and (T) does not depend of time:

(SF) $\forall_{t \leq t_{now}} \; N_t\,(F, G)$

(T) $\forall_t \; N\,(F, G)$.

Predictive power

Induction is ampliative reasoning. The conclusion of the inference is not logically entailed by the premises. The conclusion amplifies the premises. Some enumerative data is observed and a conclusion is inferred from that data. From of what have been (repeatedly) observed, we infer something concerning the observed and the unobserved. If induction is to be considered serious, then our conclusion *must* say something concerning the unobserved. Otherwise, it is not inductive reasoning. Thus, predictive power is a property of the conclusion of any inductive argument. Rejecting all propositions with predictive power, because we have no grounds for believing it, it is simply rejecting any explanation of induction.

Here is an argument for considering scientific explanatory virtues as valid in metaphysical contexts. IBE is applied in everyday, scientific and philosophical reasoning contexts. Some facts deserve explanation, in the light of competing hypothesis for explanation. Explanatory virtues are a guide to choice the best explanation of those competing hypotheses. It is controversial to say which explanatory virtues should guide our choice and it is difficult to weigh each virtue. Generally speaking, simplicity, unification, coherence or predictive power are some of those virtues. But other types of virtues can also be invoked, such as aesthetic virtues. However, is less controversial that the same virtues are valid in different contexts of reasoning. If we are naturalists, as I think we should be, we consider that there is a continuum between science and philosophy, where science is sophisticated commonsense and metaphysics is a sub-area of philosophy. To claim, *a priori*, that, for example, predictive power is not an explanatory virtue of metaphysics is a metaphysics assumption not vindicated by the successes of science. The naturalist drops this assumption.

The role of the explanatory virtue of predictive success on IBE can be illustrated by the well-known philosophical application of IBE to the realist dispute in the philosophy of science. It is worth sketching this, for it helps to flesh out my view. What best explains the predictive success of our scientific theories? Two competing philosophical explanations are available: the no-miracles explanation of Putnam and the neo-Darwinian explanation of Van Fraassen. According to Putnam (1975, 73; 1978, 18-22), the best explanation of the predictive success of our scientific theories it is their truth or approximately truth; it would be a miracle if successful scientific theories were not true or approximately true. According to Van Fraassen (1989, 39-40), the best explanation of the predictive success of our scientific theories it is a selection procedure; scientists have been selected scientific theories that have been true observed consequences.

Prima facie, in this case, both explanations are compatible. So both explanations may be inferred: scientists may select scientific theories with true observed consequences, and those theories may be true. However, if we take a look to the virtue of predictive power of the two explanations, it will be clear that the no-miracles explanation is a better explanation than is the neo-Darwinian explanation. The selective explanation does not explain why our scientific theories are not tended to be refuted by some future experience. The past true observed consequences are independent of the predictive success of the scientific theories; these consequences are a sort of an external explanation concerning the properties of the scientific theories. On the contrary, the no-miracles explanation explains why only recalcitrant experiences can falsify our best theories. Our scientific theories are true theories or approximately true theories and, hence, they go on to make successful predictions. Thus, it seems that the no-miracles explanation is a better explanation than is the neo-Darwinian explanation, because it explains the future predictive success of our scientific theories.

Now we can try to link epistemology and metaphysics. Scientific theories are true, or approximately true, because there are laws of nature. Laws of nature are the truthmakers for the scientific theories. If laws of nature are timeless necessities, then the truthmaker for the scientific theories is Omnitemporal (past, present and future exist). If laws of nature are time-limited necessities, then the truthmaker for the scientific theories is not Omnitemporal. The time-limited necessitarian is fixed to a sort of *Pastism* (only past and present truthmakers exist) concerning truthmakers and this is problematic, since this view excludes truths about the future (e.g. Armstrong 2004, 145-150). Thus, it seems that the time-limited necessitarian is committed to a sort of *time-limited scientific realism* in the realist dispute of philosophy of science.

Time-limited scientific realism is one more undesirable conception licensed by time-limited necessity. Now, a (new) explanation for theory change is available: our best scientific theories may be falsified because the proper nature of things changes. Looking into the past, for example, the time-limited scientific realist may explain the transition from phlogiston's theory to oxygen's theory claiming that, in the eighteen century, the necessary connection between *combustionness* and *phlogistonness* stopped; in the eighteen century, *combustionness* started to be connected with *oxygenness*; thus, in the eighteen century, a new theory of combustion was discover that is more faithful to our mutable nature.

6. IBE

If we connect the IBE's reasoning, the problem of induction and Beebee's (no-) solution to it, we obtain the argument:

For $t = t_a$

(1) The surprising fact: all *i*-observed *F*s are *G*s.

(2) But if $N_{t_a}(F, G)$ were true, all *i*-observed *F*s are *G*s, would be a matter of course.

(3) No available competing hypothesis can explain all *i*-observed *F*s are *G*s as well as $N_{t_a}(F, G)$ does.

($\therefore$) $N_{t_a}(F, G)$ is true.

Where $N_{t_a}(F, G)$ is (SF).

Let us suppose that at $t = t_b$, (where $t_b > t_a$) it is observed one *i*+1 *FG*. Thus, a new entity is added to the set of *i*-observed *FG*s. In order to explain this new set of *i*+1-observed *FG*s, another IBE is necessary to be made. That is, $N_{t_b}(F, G)$ must be inferred, since our background knowledge only explains why *i*-observed *F*s are *G*s. In other words, for the time-limited necessitarian, every time that a new (black) raven is observed, that observation will be waiting for an explanation.

Something very unusual is going on here. Clearly, IBE does not proceed by this way. The contrast with the timeless necessitarian is evident. Suppose that the timeless necessitarian established, via IBE, that all ravens are black. Meanwhile, he observes a new black raven. This is not a fact which deserves explanation. For him, it would be a surprise to see, for example, an albino raven, but there is no surprise to see another black raven. Instead of $N_{t_a}(F, G)$, the timeless necessitarian inferred $N(F, G)$. $N(F, G)$ is independent of time. With $N(F, G)$ in hand, and since $N(F, G) \rightarrow \forall x\,(Fx \rightarrow Gx)$, the timeless necessitarian explains *deductively* why all other observed ravens are black: *a* is *F*; $\forall x\,(Fx \rightarrow Gx)$; thus, *a* is *G*. So he is not struck by new *FG*s. On the contrary, the time-limited necessitarian cannot take the deductive step.

There is also an additional problem concerning premise (3) of the above argument. Beebee considers that (T) (i.e. $N(F, G)$), can be one of the possible truthmakers for (SF) (i.e. $N_{t_a}(F, G)$). However, as Bird & Ladyman 2011 remarks, if (T) is one of the possible truthmakers for (SF), then (T) and (SF) are not competing hypotheses. Lipton 2004 argues that if two hypotheses are not competing hypotheses, then it is impossible to make an IBE between them. For example, my car does not work. What is the best explanation available? (1) The car is out of gas; (2) the car's battery discharged *or* the car is out of gas. The two explanations are compatible. One explanation doe not pre-empts the other. Both can be true. So does not make sense, without further evidence, to make an

IBE between these two explanations.

Finally, Lipton also distinguishes between two senses in which an explanation can be the best explanation of competing plausible explanations: the *likeliest* and the *loveliest* explanations. The likeliest explanation is the 'most probable explanation'; the loveliest explanation 'as the one which would, if correct, be the most explanatory or provide the most understanding' (Lipton 2004, 59). With this distinction in hand, two versions of IBE are established: inference to the *likeliest* explanation and inference to the *loveliest* explanation. Lipton argues that we should follow the second type of inference. Inference to the likeliest explanation is trivial, because '[w]e want a model of inductive inference to describe what principles we use to judge one inference more likely than another, so to say that we infer the likeliest explanation is not helpful' (Lipton 2004, 60). In order to provide understanding, more needs to be said than pointing out to the likeliest explanation. In our case, it seems that (T) is a more loveliest explanation that is (SF).

REFERENCES:

Armstrong, D. 1983, *What is a Law of Nature?*, Cambridge, Cambridge University Press.
Armstrong, D. 2004, *Truth and Truthmakers*, New York, Cambridge University Press.
Beebee, H. 2011, "Necessary Connections and the Problem of Induction", *Nôus*, vol. 45 (3), 504-527.
Bird, A. & Ladyman, J. 2011, "Necessity Versus Regularity on Laws and the Problem of Induction", *unpublished manuscript*.
Goodman, N. 1954, *Fact, Fiction and Forecast*, Atlantic Highlands, N. J., The Athlone Press.
Lipton, P. 2004, *Inference to the Best Explanation*, London, Routledge.
Putnam, H. 1975, *Mathematics, Matter, and Method*, Cambridge, Cambridge University Press.
Putnam, H. 1978, *Meaning and the Moral Sciences*, London, Hutchinson.
Van Fraassen, B. 1989, *Laws and Symmetry*, Oxford, Oxford University Press.

A VEROSIMILITUDINARIAN ANALYSIS OF THE LINDA PARADOX

G. CEVOLANI, V. CRUPI, R. FESTA
U. of Bologna, U. of Turin, U. of Trieste

ABSTRACT: The Linda paradox is a key topic in current debates on the rationality of human reasoning and its limitations. We present a novel analysis of this paradox, based on the notion of verisimilitude as studied in the philosophy of science. The comparison with an alternative analysis based on probabilistic confirmation suggests how to overcome some problems of our account by introducing an adequately defined notion of *verisimilitudinarian* confirmation.

Key words: Conjunction fallacy, Linda paradox, Verisimilitude, Truthlikeness, Probability, Confirmation, Verisimilitudinarian confirmation.

The notion of verisimilitude or truthlikeness of a scientific theory was introduced by Karl Popper, who claimed that the main epistemic goal of science is truth approximation and that scientific progress consists in devising new theories which are closer to the truth than preceding ones (Popper, 1963). It was then extensively explored in post-Popperian theories of verisimilitude (Niiniluoto, 1987; Kuipers, 2000; Oddie, 2008). In what follows, we further pursue this line of research, by showing how a *verisimilitudinarian* analysis can be fruitfully applied also to problems emerging at the interface between epistemology and the cognitive sciences. More precisely, we focus on the so called *Linda paradox*, a key topic in current debates on the rationality of human reasoning and its limitations. In section 1, we present the Linda paradox and briefly survey the different attempts, made by both psychologists and epistemologists, to provide a satisfactory account of this phenomenon. Then, in section 2, we propose a new account based on the notion of (expected) verisimilitude, which generalizes a previous attempt in the same direction (Cevolani et al., 2010). Finally, in section 3, we compare our account with an alternative one based on Bayesian confirmation theory (Crupi et al., 2008; Tentori et al., 2012) and define a notion of *verisimilitudinarian confirmation* allowing to overcome some of the limitations of the former.

1 The Linda paradox

In a seminal work on the psychology of reasoning and judgment under uncertainty, Amos Tversky and Daniel Kahneman presented the following description of a fictitious character, Linda, which would

then become famous (Tversky & Kahneman, 1983, 297):

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

In a series of experimental inquiries, Tversky and Kahneman asked several samples of participants (both statistically naïve and sophisticated subjects) to judge the probability of some hypotheses about Linda, including the isolated statement "Linda is a bank teller" (b from now on) and the conjunctive statement "Linda is a bank teller and is active in the feminist movement" ($b \wedge f$). The results showed a strong tendency to judge $b \wedge f$ as more probable than b. In a particularly neat demonstration of the phenomenon, 142 university students were simply asked to choose the more probable state of affairs between b and $b \wedge f$: 85% of them chose the latter. This pattern of judgments is puzzling in that it conflicts with a basic and uncontroversial principle of probability theory, known as the "conjunction rule", prescribing that a conjunction of statements can not be more probable than any of its conjuncts. This "Linda paradox" is only an instance of a widespread and well documented phenomenon, usually known in the literature as the "conjunction fallacy" or the "conjunction effect". Indeed, Tversky and Kahneman themselves, along with many others in subsequent investigations, were able to replicate this phenomenon in a variety of experimental scenarios, including real-life cases like, for instance, examples of medical prognosis (Tversky & Kahneman, 1983, 301).

In the last decades, scholars interested in the analysis of human reasoning and decision making under uncertainty devoted a great deal of attention to the Linda paradox (Hertwig & Gigerenzer, 1999; Samuels et al., 2002). However, the attempt of providing a satisfactory account of the phenomenon has proved rather challenging. The psychological literature has mainly focused on whether and how the effect is modulated by several variants of the experimental task (see, for example, (Wedell & Moro, 2008)). Interestingly, ever since Isaac Levi's insightful review of Kahneman, Slovic, and Tversky's influential work (1982), the Linda paradox has attracted the attention of a number of epistemology scholars (Levi, 1985; Bovens & Hartmann, 2003; Hintikka, 2004; Crupi et al., 2008; Peijnenburg et al., 2012). Their accounts can all be seen as attempts to flesh out the otherwise esoteric statement by Tversky and Kahneman themselves that "*feminist bank teller* is a better hypothesis about Linda than *bank teller*" (Tversky & Kahneman, 1983, 311). In what follows, we will explore a different strategy to account for this noteworthy remark by providing a verisimilitudinarian analysis of the problem. In a nutshell, we will show that "feminist bank teller", while less likely to be *true* than "bank teller", may well be more likely to be a better approximation to the *whole truth* about Linda.

2 A verisimilitudinarian account of the Linda paradox

2.1 Verisimilitude vs probability

In general terms, a statement, hypothesis or theory h is highly verisimilar if it says many things about the target domain, and if many of these things are (almost exactly) true. Thus, the (degree of) verisimilitude of h must depend on both its *content*, i.e., how much h says, and its *accuracy*, i.e., how much of what h says is in fact true. In Popper's words, verisimilitude "represents the idea of approaching comprehensive truth. It thus combines truth and content" (Popper, 1963, p. 237).

An important consequence of the fact that verisimilitude is a combination of truth and content — or a "mixture of truth and information" (Oddie, 1986, 12) — is that the notions of verisimilitude and probability have to be carefully distinguished. In particular, it is possible that h is expected to be quite close to the whole truth about the domain, but still h is not expected to be true, i.e., h is not probable. In fact, while (conditional) probability is a *decreasing* function of content, (expected) verisimilitude must be *positively* associated to high content. This is simply because "nothing is as close to the truth as the whole truth itself" (Oddie, 1986, 11), the latter clearly being a uniquely accurate *and exhaustive* description of a given matter of interest.[1]

This last remarks suggests how a verisimilitudinarian account of the Linda paradox can work. The basic idea is that experimental participants may judge $b \wedge f$ a better hypothesis about Linda as compared to b because they evaluate $b \wedge f$ as more verisimilar than b (Cevolani et al., 2010). In fact, the hypothesis "feminist bank teller", while less likely to be true than "bank teller", may well be evaluated as a better approximation to the whole truth about Linda. In this connection, it is perhaps surprising that — even without mentioning the idea of verisimilitude, of which they were probably unaware — Tversky and Kahneman themselves anticipated the basic intuition underlying a verisimilitudinarian analysis of the Linda paradox, as the following telling quotation reveals (Tversky & Kahneman, 1983, 312):

> The expected value of a message can sometimes be improved by increasing its content, although its probability is thereby reduced. [...] Consider the task of ranking possible answers to the question "What do you think Linda is up to these days?" The maxim of value could justify a preference for $b \wedge f$ over b in this task, because the added attribute *feminist* considerably enriches the description of Linda's current activities at an acceptable cost in probable truth.

[1] More precisely, if h logically entails g, but g does not entail h, then h is less probable, but may well be more verisimilar, than g. Indeed, if h and g are both true, and h is logically stronger than g, then h is more verisimilar than g according to most verisimilitude theorists (an exception being (Oddie, 1986)).

In other words, we suggest that experimental participants are assessing the expected verisimilitude, not the probability, of the two hypotheses involved in the Linda problem, and that this can explain their preference for $b \wedge f$ over b.

2.2 Expected verisimilitude of conjunctive hypotheses

To illustrate our analysis, we introduce a general framework to address the Linda paradox, based on the so-called basic feature approach to verisimilitude (Cevolani et al., 2011). This approach focuses on *conjunctive hypotheses*, i.e., hypotheses expressed as conjunctions of logically independent atomic statements of a language $\mathscr{L}$ describing a target domain $\mathscr{U}$. The (degree of) verisimilitude $Vs(h)$ of a conjunctive hypothesis h expresses the closeness or similarity of h to "the whole truth" about $\mathscr{U}$, construed as the most complete true description of the basic features of $\mathscr{U}$ in $\mathscr{L}$. Since, in most interesting cases, the whole truth is simply unknown, the estimated verisimilitude of competing hypotheses, not their actual verisimilitude, is the crucial point of interest. Accordingly, the (degree of) expected verisimilitude $EVs(h|e)$ of h on e expresses its estimated degree of closeness to the truth with respect to the available evidence e (Oddie, 1986; Niiniluoto, 1987). Given an epistemic probability distribution $p(\cdot)$ over the set of the relevant alternative states of affairs (possible worlds) $c_1, \ldots, c_n$ and a measure $Vs(h, c_i)$ of the closeness of h to a given state of affairs c_i, $EVs(h)$ is defined as follows:

$$EVs(h|e) = \sum_{c_i} Vs(h, c_i) \times P(c_i|e) \tag{1}$$

We suggest that, in order to account for the Linda paradox, one should focus on the expected verisimilitude of the two (conjunctive) hypotheses involved — i.e., on $EVs(b|e)$ and $EVs(b \wedge f|e)$, where e is Linda's description—, and on their relationship.

Verisimilitude theorists have proposed a number of (partially conflicting) measures Vs of verisimilitude. Most of these measures, however, agree with each other as far as the verisimilitude of a conjunctive hypotheses h is concerned. In particular, most of them are *monotonic* in the sense that the verisimilitude of h increases with the addition of true conjuncts, and decreases with the addition of false conjuncts (Cevolani et al., 2011, 187). As an example, a monotonic verisimilitude measure Vs would order the two hypotheses involved in the Linda scenario as follows:

$$\begin{aligned} &\text{if } f \text{ is true in } c_i\text{, then} \quad Vs(b \wedge f, c_i) > Vs(b, c_i) \\ &\text{if } f \text{ is false in } c_i\text{, then} \quad Vs(b \wedge f, c_i) < Vs(b, c_i) \end{aligned} \tag{2}$$

One can then prove that, under suitably defined conditions, for all monotonic verisimilitude measures

Vs:[2]

$$EVs(b \wedge f|e) > EVs(b|e) \text{ iff } P(f|e) > \sigma \tag{3}$$

where σ is a threshold value depending on the specific measure *Vs*. In words, if "Linda is a feminist" is sufficiently likely to be true given Linda's story (i.e., $P(f|e) > \sigma$), "feminist bank teller" is estimated as more verisimilar than "bank teller" alone. This makes the former a better hypothesis about Linda than the latter, thus explaining both people's preferences and Tversky and Kahneman's "esoteric" remark quoted in section 1.

3 Verisimilitude and confirmation

According to the analysis presented in the previous section, people tend to prefer "feminist bank teller" ($b \wedge f$) over "bank teller" alone (b) because they judge f sufficiently probable given Linda's story (e) and hence evaluate $b \wedge f$ as a better approximation to the whole truth about Linda than b.

As one can see from theorem 3, the relatively high probability of f given e is considered as the main factor underlying the Linda paradox. This is a point that the present analysis has in common with most alternative accounts of this phenomenon. However, this aspect of the verisimilitudinarian account is problematic, as the following case shows (Tentori et al., 2012). Suppose that a new item of information is added in the Linda scenario, i.e., the statement that "Linda owns an umbrella" (u). It seems clear that u is extremely likely to be true (since almost everybody owns an umbrella) and at least as probable as f, even on the background assumption that e. Accordingly, by the same pattern of reasoning described above, participants should prefer $b \wedge u$ over b alone as an hypothesis about Linda. In fact, one can see from theorem 3 that $b \wedge u$ is estimated as more verisimilar than b when $P(u) > \sigma$. In turn, since we are assuming that $P(u) \geq P(f)$, inequality $P(u) > \sigma$ holds whenever $P(f) > \sigma$. It follows that participants should prefer $b \wedge u$ to b whenever they prefer $b \wedge f$ to b.

The problem with this account is that, although f and u are both highly probable given e, only the former is *confirmed* by e, in the sense that its probability is increased by e. On the contrary, we can assume that the probability of u remains unchanged after learning e. Indeed, recent experimental investigations (Tentori et al., 2012) show that, when u is added in the Linda scenario, participants are much less prone to the conjunction fallacy; more precisely, the number of subjects preferring $b \wedge u$ to b is much lesser than the number of those judging $b \wedge f$ a better hypothesis than b about Linda.

A possible explanation of these results is that the key variable that generates the effect is not the high *posterior probability* of the added conjunct (like f or u), but the (*probabilistic*) *confirmation*

[2]Proofs are omitted through the paper; for details, see (Cevolani et al., 2012).

that the added conjunct receives from Linda's story. In general terms, the (degree of) confirmation $c(h,e)$ given by evidence e to hypothesis h expresses how much the probability of h increases (or decreases) once e is learned (Festa, 1999; Crupi & Tentori, 2012). As an example, the well-known "difference measure" of confirmation is defined as follows (Carnap, 1962, 361):

$$c(h,e) = p(h|e) - p(h) \tag{4}$$

Note that $c(h,e)$ is positive if and only if $P(h|e) > P(h)$, i.e., if the probability of h is increased by e. According to the "confirmation account" of the Linda paradox (Crupi et al., 2008), participants are not assessing the probability of the hypotheses at issue, but their confirmation given Linda's story. Under suitably defined conditions, this account can explain why participants prefer $b \wedge f$, but not $b \wedge u$, to b as an hypothesis about Linda. Indeed, this preference might be due to the fact that $c(b \wedge f, e) > c(b,e)$ while $c(b \wedge u, e) \leq c(b,e)$.

The intuitions underlying the confirmation account presented above and the verisimilitudinarian account of section 2 can be interestingly combined in what might be called the *verisimilitudinarian confirmation* account (*vs-confirmation* account for short) of the Linda paradox. The basic idea is that people might be assessing neither the verisimilitude, nor the confirmation, of the hypotheses involved in the Linda problem, but their vs-confirmation, which can be defined for instance as follows (cf. definition 4): given a verisimilitude measure Vs,

$$c_{Vs}(h,e) = EVs(h|e) - EVs(h) \tag{5}$$

Note that $c_{Vs}(h,e)$ is positive if and only if $EVs(h|e) > EVs(h)$; in other words, $c_{Vs}(h,e)$ expresses how much the expected verisimilitude of hypothesis h increases (or decreases) on the basis of evidence e.

According to the vs-confirmation account, "feminist bank teller" would be preferred to "bank teller" alone just in case $c_{Vs}(b \wedge f, e)$ were greater than $c_{Vs}(b,e)$, i.e., whenever $EVs(b \wedge f)$ is increased by e more than $EVs(b)$. One obtains a particularly interesting application of this account if definition 5 of vs-confirmation is phrased in terms of so called *additive* verisimilitude measures Vs. A measure Vs is additive when it is monotonic and, given a conjunctive hypothesis $h = h_1 \wedge \cdots \wedge h_k$ (where h_i are logically independent atomic statements), $Vs(h) = \sum_{h_i} Vs(h_i)$. Thus, for instance, if Vs is additive

then $Vs(b \wedge f) = Vs(b) + Vs(f)$.[3] One can then prove that, for all additive verisimilitude measures Vs,

$$c_{Vs}(b \wedge x, e) \gtreqless c_{Vs}(b, e) \text{ iff } c(x, e) \gtreqless 0 \tag{6}$$

i.e., that $b \wedge x$ is vs-confirmed by e more than b just in case the added conjunct x is (probabilistically) confirmed by e. An attractive feature of the vs-confirmation analysis of the Linda paradox is that it can easily accommodate the umbrella case. In fact, such case was originally introduced in order to suggest that f is confirmed by e while u is not, i.e., that the following inequalities hold: $c(f, e) > 0 \geq c(u, e)$. By theorem 6, this implies that $b \wedge f$, but not $b \wedge u$, is more vs-confirmed by e than b, i.e., that $c_{Vs}(b \wedge f, e) > c_{Vs}(b, e) \geq c_{Vs}(b \wedge u, e)$. In turn, this provides an explanation for the participants' preferences based on the notion of verisimilitudinarian confirmation.

Acknowledgements

Gustavo Cevolani and Vincenzo Crupi acknowledge financial support by Grant CR 409/1-1 to Vincenzo Crupi from the *Deutsche Forschungsgemeinschaft* (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516).

REFERENCES

Bovens, L. & Hartmann, S. (2003). *Bayesian epistemology*. Oxford: Oxford University Press.

Carnap, R. (1962). *Logical Foundations of Probability*. Chicago: University of Chicago Press. First edition: 1950.

Cevolani, G., Crupi, V., & Festa, R. (2010). The whole truth about Linda: probability, verisimilitude and a paradox of conjunction. In M. D'Agostino, G. Giorello, F. Laudisa, T. Pievani, & C. Sinigaglia (Eds.), *SILFS New Essays in Logic and Philosophy of Science* (pp. 603–615). London: College Publications.

Cevolani, G., Crupi, V., & Festa, R. (2011). Verisimilitude and belief change for conjunctive theories. *Erkenntnis*, *75*(2), 183–202.

Cevolani, G., Crupi, V., & Festa, R. (2012). A novel analysis of the linda paradox. Manuscript.

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation and the conjunction fallacy. *Thinking and Reasoning*, *14*, 182–199.

Crupi, V. & Tentori, K. (2012). Confirmation theory. In A. Hájek & C. Hitchcock (Eds.), *The Oxford Handbook of Philosophy and Probability*. Oxford: Oxford University Press. Forthcoming.

Festa, R. (1999). Bayesian confirmation. In M. C. Galavotti & A. Pagnini (Eds.), *Experience, Reality, and Scientific Explanation*. Dordrecht: Kluwer Academic Publishers.

Hertwig, R. & Gigerenzer, G. (1999). The 'conjunction fallacy' revised: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12*, 275–305.

Hintikka, J. (2004). A fallacious fallacy? *Synthese*, *140*, 25–35.

[3] A family of additive verisimilitude measure is provided by so called *feature contrast measures* of verisimilitude (Cevolani et al., 2011).

Kuipers, T. A. F. (2000). *From Instrumentalism to Constructive Realism*. Dordrecht: Kluwer Academic Publishers.

Levi, I. (1985). Illusions about uncertainty. *British Journal for the Philosophy of Science*, *36*, 331–340.

Niiniluoto, I. (1987). *Truthlikeness*. Dordrecht: Reidel.

Oddie, G. (1986). *Likeness to truth*. Dordrecht: Reidel.

Oddie, G. (2008). Truthlikeness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Peijnenburg, J., Fitelson, B., & Douven, I. (Eds.). (2012). *Special issue on "Probability, Confirmation and Fallacies"*. *Synthese*, 184.1.

Popper, K. R. (1963). *Conjectures and Refutations*. London: Routledge and Kegan Paul.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (Ed.), *Common Sense, Reasoning and Rationality* (pp. 236–268). New York: Oxford University Press.

Tentori, K., Crupi, V., & Russo, S. (2012). On the determinants of the conjunction fallacy: Probability vs. inductive confirmation. *Journal of Experimental Psychology: General*. Forthcoming.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293–315.

Wedell, D. H. & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: effects of response mode, conceptual focus, and problem type. *Cognition*, *107*(1), 105–136.

UNDERDETERMINATION, REALISM, AND TRUTH APPROXIMATION

GUSTAVO CEVOLANI [1], LUCA TAMBOLO [2]
[1] Department of Philosophy, University of Bologna
g.cevolani@gmail.com
[2] Department of Humanistic Studies, University of Trieste
l_tambolo@hotmail.com

ABSTRACT: Held (2011) attacks scientific realism via a criticism of the No-Miracles Argument based on the underdetermination of theories. In this paper we argue that the No-Miracles Argument, when deployed in conjunction with non-naive versions of realism, based on the view that verisimilitude or truth approximation is the main cognitive aim of scientific inquiry, survives Held's criticism unscathed.

Keywords: No-Miracles Argument, underdetermination, realism, anti-realism, success, verisimilitude, truthlikeness, truth approximation.

Introduction

The so-called 'No-Miracles Argument' (NMA) for scientific realism is the thesis that the success enjoyed by scientific inquiry would be an inexplicable fact, if scientific theories did not provide (at least approximately) true descriptions of (some of the relevant features of) the world. In this regard, Hilary Putnam famously claimed that scientific realism is 'the only philosophy that doesn't make the success of science a miracle' (1975, 73). Underdetermination (UD) may be taken as the 'simple, incontestable fact that observational evidence is logically inconclusive with respect to the truth of [a] theory' (Leplin 2004, 119). Based on UD, a whole family of arguments, coming in varying degrees of scope and strength, have been put forward against realism and in favour of anti-realism. The fundamental insight underlying all these arguments is that the realist's inference from the success of a theory T to the (approximate) truth of T cannot be granted, given UD, on purely logical grounds, and so requires additional justification of some kind, which may be difficult if not impossible to provide (Laudan 1981; Niiniluoto 1999).

The recession-proof literature on the issue of UD and scientific realism has recently been enriched by an exchange between Held (2011; 2012) and Morganti (2011; 2012). In his 'Truth does not explain predictive success' (2011), Held attacks realism by arguing that NMA is easily refuted, when the consequences of UD are taken into account. Since NMA is sometimes referred to as no less than 'the ultimate argument for scientific realism' (after van Fraassen 1980, 39), a reply on the part of the realists is urgently needed. In this paper we shall neglect the nuances of the debate

between Held and Morganti, in which, it seems to us, some crucial points are missed. Rather, after critically reconstructing Held's argument against NMA (section 2), we shall argue that the so-called 'verisimilitudinarian' account of scientific inquiry (henceforth: VS) provides the means to rebut Held's criticism (section 3). Revolving around the idea that the main cognitive aim of science is the acceptance of theories with a high degree of verisimilitude (a.k.a. truthlikeness, or approximation, or closeness, to the whole truth), VS allows one to conclude (section 4) that Held's criticism of NMA may hurt, at best, a naive version of realism, based on the view that the main goal of scientific inquiry is the search for plainly true theories.

1 . Success, truth, and verisimilitude

1.1. SUCCESS, THE NO-MIRACLES ARGUMENT, AND UNDERDETERMINATION

As Held puts it, NMA is the claim that the best explanation of the success of a scientific theory *T* is provided by its truth.[1] According to Held, a straightforward refutation of NMA is readily available. Such a refutation is based on UD, i.e., on the fact that 'observational evidence is logically inconclusive with respect to the truth of [a] theory' (Leplin 2004, 119). This implies that, as both realists and antirealists agree, observational evidence underdetermines theories, in the sense that the truth of a theory *T* cannot be deduced from its success. From this, it immediately follows the possibility that a successful theory is false. This fact alone is sufficient, Held argues, to show that a theory's truth 'cannot be an explanation of its success' (2011, 234). Therefore, NMA is refuted: an explanation of a theory's success 'must be compatible with the assumption that this theory is false' (2011, 234).

At first glance, one may get the impression that Held is making an obviously false claim (i.e., that UD logically entails the negation of NMA), and that he is therefore attributing to scientific realists the view that the success-to-truth link must be without exceptions, i.e., a deductive one. However, this cannot be the case, since Held (2011, 233) seems ready to accept that NMA is an abductive, i.e., a non-deductive, inference. In other words, this means that NMA does not amount to the claim that the success of a theory *T* entails its truth. Rather, NMA is the thesis that the success of *T* is a *fallible indicator* of its truth – indeed, this is how all 'critical' or 'fallibilist' realists (such as,

1

 One can distinguish (see. e.g., Niiniluoto 1999, 164-165) between empirical success (which obtains in case a theory *T* entails, in conjunction with some initial conditions, true empirical statements) and predictive success (which obtains in case a theory *T* entails, in conjunction with some initial conditions, true empirical statements about the future), the latter being a special case of the former. Although Held's criticism of NMA is phrased in terms of predictive success, in this paper we shall speak, in more general terms, of 'success', meaning 'empirical success', since it is the success of science in general, and not only its predictive success, that NMA is aimed at explaining.

for instance, Niiniluoto 1999; Psillos 1999; Kuipers 2000; and many others) understand NMA. Moreover, all non-naive realists readily acknowledge that highly successful theories are typically, strictly speaking, false – for instance, because they are idealized. As a consequence, they typically maintain that the best explanation of the success of a theory *T* is provided by its approximate truth, or by its verisimilitude (or truthlikeness), and not by its truth *simpliciter*. Therefore, we argue, NMA is not threatened by the consequences of UD (i.e., the possibility of the existence of successful false theories) that Held introduces as refuting evidence against it – at least, not when NMA is deployed in conjunction with non-naive versions of realism.

1.2. SUCCESS AS A FALLIBLE INDICATOR OF (CLOSENESS TO THE) TRUTH

In any case, it seems to us that Held is trying to make a more interesting point. Suppose that *T* is successful w.r.t. evidence *E*, in the sense that *T* entails *E*, and suppose that *E* is true. This means that our actual world, *w*, is an *E*-world, i.e., one of the possible worlds where *E* is true (Held 2011, 234). The realist suggests that, given the success of *T*, *w* either is a (*T* & *E*)-world or is a (¬*T* & *E*)-world, which however is close to a (*T* & *E*)-world; in other words, that *T* is true or is at least approximately true. Thus, the realist owes us an argument to the effect that we have some reason to think that *w* is actually in the neighbourhood of *T*. A graphical representation of the situation may help in clarifying the problem. In Figure 1, both evidence *E* and theory *T* are represented as the corresponding sets of possible worlds which are models of *E* and *T*, respectively.

Figure 1: Both *T* and *T'* are false successful theories, but *T* is a better approximation than *T'* to the whole truth represented by the actual world *w*.

Since we are assuming that *E* is true, the actual world, *w*, can be represented as a dot located within the *E*-area (in gray). Moreover, since *T* is successful, we also know that *T* is somewhere within *E* – although we do not know exactly where: in any case, it can be represented by the *T*-circle including all the *T* & *E*-worlds. Due to UD, the area between *T* and *E*, which includes the (¬*T* & *E*)-worlds, is (densely) populated by other theories which are also successful, because they entail

E. Each one of these theories can be represented as a circle within *E*; for simplicity, we shall consider only one of these theories, *T′*.

Held's challenge to the realist can be now rephrased as follows: how do we know that *T* is at least close to *w*, and not far away near the edge of *E*, like *T′*? In other words, how do we know that *T*, besides not being true, is not badly or seriously false, i.e., very distant from the truth?[2] Here, Held (2011, 234) seems to think that the only possible answer for the realist would be to show that, were *T* not true, *T* would not be successful; in other words, that were *w* not within the *T*-area, then the *T*-area would be outside the *E*-area). But this would mean, by contraposition, that the success of *T* logically entails its truth – and as we said above, no non-naive realist would be prepared to embrace such a line of argument. Therefore, Held concludes, the realist is left with no good argument to the effect that *T* is successful because of its approximate truth – and therefore, NMA is refuted.

2 . The verisimilitudinarian approach to scientific inquiry

2.1. VERISIMILITUDE AS APPROXIMATION TO THE TRUTH

As mentioned above, it seems to us that the verisimilitudinarian approach to scientific inquiry (VS) provides realists with the tools to meet Held's challenge. The notion of verisimilitude or truthlikeness of theory *T* was introduced by Popper (1963) and then further investigated by such authors as Miller (1978; 2006), Oddie (1986, 2011), Kuipers (1987; 2000), Niiniluoto (1984; 1987; 1999), and Schurz and Weingartner (1987; 2010).[3]

In informal terms, *T* is highly verisimilar if it says many things about the target domain, and if many of these things are (almost exactly) true. Thus, the (degree of) verisimilitude of *T* must depend on both its content, i.e., how much the theory says, and its accuracy, i.e., how much of what the theory says is in fact true. In Popper's words, verisimilitude 'represents the idea of *approaching comprehensive truth*. It thus combines truth and content' (Popper 1963, 237, emphasis added). With reference to Figure 1, where the actual world *w* represents 'the (whole) truth' about the domain,

2

 This is our best attempt to understand Held's talk about successful theories which are 'totally' or 'entirely' false (2011, 233). These expressions, we assume, denote successful theories with a very low degree of verisimilitude or of approximate truth. In this connection, one may note that, if *T* is logically closed, then *T* cannot be 'entirely false' in the sense of entailing only false statements, since all logically closed theories contain at least all the tautologies (and typically also other factually true statements).

3

 For a survey of the history of VS see Niiniluoto (1998) and, for an informal overview, Cevolani and Tambolo (2012).

theory *T* is highly verisimilar because it is both informative (the *T*-circle is 'small') and close to the truth; instead, the verisimilitude of *T′* is much lower, since *T′* is both less informative and farther from the truth than *T*. To sum up, verisimilitude is a 'mixture' (Oddie 1986, 12) of two ingredients, truth and content. If truth were the only ingredient, then all truths, including the tautology, would be equally (and maximally) verisimilar; and, vice versa, if only content were relevant, then a plain contradiction would be closer to the truth than any other theory. Thus, devising highly verisimilar theories is a 'game of excluding falsity and preserving truth' (Niiniluoto 1999, 73).

On the basis of this broad characterization of verisimilitude, the key tenets of VS can be briefly formulated as follows: (1) since some *false* theories (like *T* in Figure 1) are closer to the truth than others (like *T'*), a false but highly verisimilar theory can constitute a genuine approximation to the truth, and hence an instance of progress, when adopted to replace a less verisimilar theory;[4] (2) we can devise methods to fallibly assess, on the basis of the available evidence, which among two competing theories is closer to the truth. This latter tenet is of crucial importance, since, in most interesting cases, the truth is simply unknown, so that the *estimated verisimilitude* of competing theories, not their verisimilitude, is the real point of interest. Accordingly, the theory of verisimilitude has to address both a logical and an epistemic problem. The *logical* problem of verisimilitude amounts to the preliminary definition of an appropriate notion of verisimilitude, allowing for a comparison of any two theories with regard to their closeness to the, supposedly known, truth. The *epistemic* problem of verisimilitude, on the other hand, amounts to the definition of an appropriate notion of estimated verisimilitude by which the estimated closeness to the unknown truth of any two theories could be compared on the basis of the available evidence.

2.2. ESTIMATING VERISIMILITUDE

Verisimilitude theorists have proposed different methods to deal with the issue of estimated verisimilitude.[5] For instance, within the context of his 'critical scientific realism', Niiniluoto (1987; 1999) defines a quantitative notion of 'expected verisimilitude' in Bayesian terms, providing a measure EVs($T|E$) of the degree of estimated closeness to the truth of any theory *T* given the

4

For instance, such theory-changes as that from Kepler's to Newton's or from Newton's to Einstein's theory are progressive because, although the new theory is, strictly speaking, (presumably) false, we have good reasons to believe that it is closer to the truth than the superseded one.

5

See, in particular, Oddie (1986, Ch. 7), Niiniluoto (1987, Ch. 7), and Kuipers (1987, Chs. 7 and 9). Note that some verisimilitude theorists (like Miller 2006, Chs. 10-11) are skeptic about most of these answers to the epistemic problem of verisimilitude.

evidence *E*. This definition is based both on a suitably defined measure of verisimilitude *Vs* and on an underlying epistemic probability distribution *P* over the set of the relevant alternative states of affairs or possible worlds w_1, …, w_m (cf. Figure 1). Niiniluoto proposes a number of methods to measure the degree of verisimilitude $\mathrm{Vs}(T,w_i)$, expressing the closeness or similarity of a theory *T* to a given state of affairs w_i. Degrees of probability $\mathrm{P}(w_i|E)$ express instead rational degrees of belief in the truth of each alternative state, given the evidence. The expected degree of verisimilitude of *T* is then calculated by summing up the verisimilitude of *T* in each state multiplied by the corresponding probability (Niiniluoto 1987, 269):

$$\mathrm{EVsTE}=\mathrm{wiVs(T,wi)}\times\mathrm{P(wi|E)}$$

If, as VS assumes, verisimilitude is the main cognitive goal of science, then the choice between competing theories should be based on their expected verisimilitude, construed as the relevant 'epistemic utility' guiding the acceptance of scientific hypotheses. In Niiniluoto's approach, this idea can be explicitly formulated in the form of the following 'rule of acceptance' (1987, 416): 'Accept on evidence *E* that theory *T* which maximizes the value $\mathrm{EVs}(T|E)$'. Thus, the notion of estimated verisimilitude allows us to make fallible but educated guesses both about the relative verisimilitude of competing theories and about what theory should be regarded as the most verisimilar one on the basis of available evidence.

3. Toward a verisimilitudinarian version of NMA

Within VS, the (whole) truth is seen as the ideal goal of inquiry. Nevertheless, it seems extremely unlikely that we shall ever be able to reach such a goal, since as far as we know even our best theories may well be false. As a consequence, the main cognitive goal of scientific research is assumed to be the search for highly verisimilar theories, i.e., theories which, though presumably false, are close to the truth. (Notice that the maximally verisimilar theory is the truth itself; thus, aiming at highly verisimilar theories also means to aim at the truth seen as the ideal goal of inquiry.)

From the vantage point of VS, the success of science is far from being a miracle: it is a fact that can be reasonably explained by the increasing verisimilitude of our theories. According to VS, such methods for the estimation of verisimilitude as Niiniluoto's, briefly described in section 3.2, allow one to *tentatively* accept the best theory *T* at disposal, i.e., the one which is estimated as the closest to the truth given the available evidence *E*. This means that, even if we know that *T* is false, we may have good reasons to believe that *T* is approximately true or verisimilar. In turn, this justifies the (fallible and tentative) inference from the success of *T* to its closeness to the truth, that is, what we may dub a verisimilitudinarian version of NMA (cf. Niiniluoto 1999, sections 6.4-6.5).

Such a version of NMA, it seems to us, can easily escape Held's criticism, for at least two reasons. First of all, the possibility of the existence of a false successful theory is readily granted by the supporters of VS – indeed, that scientists deal most of the time with false theories is one of the main premises of VS. Moreover, on the epistemic level, a verisimilitudinarian version of NMA seems well suited to answer Held's challenge (cf. again Figure 1). In fact, even granted that the success of T does not entail its truth, and that many other theories like T' are successful with respect to the same evidence E, T may still be estimated as more verisimilar than each of its rivals on E, and then fallibly accepted as the best theory at disposal. In this precise sense, even if, as Held claims, truth does not explain success, (estimated) closeness to the truth does. Therefore, we conclude, NMA, when deployed in conjunction with non-naive versions of realism, based on the view that truth approximation is main cognitive aim of science, survives Held's criticism unscathed.

Acknowledgments

Thanks are due to an anonymous referee, whose comments helped us to clarify the structure of our argument. Gustavo Cevolani acknowledges financial support from PRIN grant 2008 'Probability, stability, and invariance'. Luca Tambolo acknowledges financial support from PRIN grant 2008 'Probability, confirmation, and verisimilitude'.

REFERENCES

Cevolani, G., and Tambolo, L. 2012, 'Progress as approximation to the truth: a defence of the verisimilitudinarian approach', *Erkenntnis* DOI: 10.1007/s10670-012-9362-y.

Held, C. 2011, 'Truth does not explain predictive success', *Analysis* 71, 232-234.

Held, C. 2012, 'A particularist defence of scientific realism? Reply to Morganti', *The Reasoner* 6.2, 22-23.

Kuipers, Th.A.F. 1987, 'A structuralist approach to truthlikeness', in Kuipers, Th.A.F. (ed.), *What is closer-to-the-truth?*, Amsterdam, Rodopi, 79-99.

Kuipers, Th.A.F. 2000, *From Instrumentalism to Constructive Realism*, Dordrecht, Kluwer.

Laudan, L. 1981, 'A confutation of convergent realism', *Philosophy of Science* 48, 19-48.

Leplin, J. 2004, 'A theory's predictive success can warrant belief in the unobservable entities it postulates', in Hitchcock, C. (ed.), *Contemporary Debates in Philosophy of Science*, Malden, Blackwell, 117-133.

Miller, D. 1978, 'On distance from the truth as a true distance', in Hintikka, J., Niiniluoto, I., and Saarinen, E. (eds.), *Essays on Mathematical and Philosophical Logic*, Dordrecht, Kluwer, 415-435.

Miller, D. 2006, *Out of Error: Further Essays on Critical Rationalism*, Aldershot, Ashgate.

Morganti, M. 2011, 'Truth and success: reply to Held', *The Reasoner* 5.7, 106-107.

Morganti, M. 2012, 'Truth and success, again: reply to Held on generalist versus particularist (anti-)realism', *The Reasoner* 6.6, 99-100.

Niiniluoto, I. 1984, *Is Science Progressive?*, Dordrecht, Reidel.

Niiniluoto, I. 1987, *Truthlikeness*, Dordrecht, Reidel.

Niiniluoto, I., 1998, 'Verisimilitude: the third period', *The British Journal for the Philosophy of Science* 49, 1-29.

Niiniluoto, I. 1999, *Critical Scientific Realism*, Oxford, Oxford University Press.

Oddie, G. 1986, *Likeness to Truth*, Reidel, Dordrecht.

Oddie, G. 2011, 'The content, consequence and likeness approaches to verisimilitude: compatibility, trivialization, and underdetermination', *Synthese*, DOI: 10.1007/s11229-011-9930-8.

Popper, K.R. 1963, *Conjectures and Refutations*, London, Routledge & Kegan Paul.

Psillos, S. 1999, *Scientific Realism. How Science Tracks Truth*, London and New York, Routledge.

Putnam, H. 1975, *Mathematics, Matter and Method*, Cambridge: Cambridge University Press.

Schurz, G. and Weingartner, P. 1987, 'Verisimilitude defined by relevant consequence-elements. A new reconstruction of Popper's original idea', in Kuipers, Th.A.F. (ed.), *What is closer-to-the-truth?*, Amsterdam, Rodopi, 47-78.

Schurz, G. and Weingartner, P. 2010, 'Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude', *Synthese* 172, 415-436.

Van Fraassen, B.C. 1980, *The Scientific Image*, Oxford, Oxford University Press.

THE FEASIBILITY OF DETERMINABLES AND ITS RELATION TO THE SCIENTIFIC IMAGE

JUAN COLOMINA
Department of Philosophy
The University of Texas at Austin
Colomina-Alminana_Juan@austin.utexas.edu

ABSTRACT: Objects have properties. Stones are solid, particles have mass, flowers have colors. All these instances exemplify how the world consists of entities possessing a property or determinate from a given category or determinable. According to this view, the world has a determinate/determinable structure. But, what does it really mean that the world has a determinate/determinable structure? The main purpose of this paper is to answer this question and explore the relation between the notions of determinate and determinable and our current scientific image of the world.

First, we shall discuss the traditional distinction between determinate and determinable. I shall support that this classical approach does not provide a satisfactory explanation of the nature of determinables. This distinction solely based on the notion of extensionality or co-determinability (which holds among attributes that belong in the same determinable) avoids the question of what exactly a determinable is. Then, we shall defend that it is necessary to reformulate the relation between determinate and determinable within a space that is restricted to relatively determinate states (or partial information scenarios), and that this fact can provide a definition of what the candidate determinates should be.
KEY WORDS: Determinate/Determinable distinction, Properties, Scientific image, Universal versus Particular, Point of views, Perspective, Intensionality, R. Carnap, K. Fine.

* * *

The distinction between determinate and determinable is introduced by W. E. Johnson in the context of logical categories that we use to classify logical objects, specifically when he discusses "the ultimate compromising classes" (Johnson 1922, 60). According to him, a determinable can be understood as a term that can have several different specifications or determinates. Somehow, determinables are adjectives (Johnson 1921, 173).

This conception relies on the traditional account of the principles of logical division where a class is represented as consisting of subclasses, or the classical distinction between *genus* and *species* (Johnson 1921, 173). This division assumes the mutual exclusivity between subclasses, the exhaustibility of a class and the existence of some principle of *fundamentum divisionis* (that is, the principle that underlies the division). In this distinction, the *fundamentum divisionis* can be understood as a determinable, since in many cases a class of things is divided according to color,

size or some other aspect according to which the particular elements of the class can be compared (Johnson 1921, 173-4).

Furthermore, Johnson claims that determinates under a determinable are in turn also adjectives, but they are more specific than the determinable. That is, a determinable can always be further specified and this further specification produces determinates (Johnson 1921, 177). According to Johnson, this kind of distinction is clearly exemplified by sentences like

(1) Red is a color.

(2) Plato is a man.

Both sentences seem to represent the same relation between a determinate and its determinable. But, (2) involves a real adjectival predication because 'man' is an adjective that can be predicated of Plato. In turn, 'color' is not an adjective that can be predicated of red (Johnson 1921, 176). The reason is simple. Given that, according to Johnson, the relationship between a determinate and its determinable resembles that of an individual to a class (Johnson 1921, xxxv), Plato can belong in many different classes; in this sense, Plato is a multidimensional entity. On the contrary, red only belongs in the class of colors and in this sense is one-dimensional.

Following this approach, the relationship between determinable and determinate is not the same as between a single object and some class to which it belongs, despite the similarity of sentences like 'Red is a color' and 'Plato is a man'. In the case of membership to a class, all the members of such a class are characterized by some adjective or combination of adjectives, whereas in the case of determinates there is no such common (secondary) adjective (Johnson 1921, 176). If this is the case, when we claim that red is a color we are only asserting a contingent fact.

An appropriate explanation should provide a necessary truth because, ultimately, what Johnson defends is that "my distinction between substantive and adjective is roughly equivalent to the more popular philosophical antithesis between particular and universal; the notions, however, do not exactly coincide" (Johnson 1922, xiii). Consequently, the nature of a determinable should not be explainable in terms of its composition or extension, as the particular determinates that belong in the determinable enjoy a contingent relativism with the latter, and this avoids the question what a determinable really is.

Johnson also seems to think that determinables and determinates under them are both predicates of objects of the same order. This means that if for example an object is red the determinable color is also predicable of it. But how to understand this predication remains open in his presentation.

A more problematic issue is that Johnson seems to consider a conjunctive relation between determinable and determinate. If we suppose that between a determinable and its determinates exists a relation of specificity or subordination, we can only conclude that this position defends a mere unilateral implication between them. For instance, if *x* is a dog, we can say that *x* necessarily is an animal (because 'dog' is a subordinate determinate of the determinable 'animal') but we can never say that an animal is necessarily a dog. If we assume a conjunctive relationship between them, then the relation between determinate and determinable is asymmetric. Furthermore, if we accept Johnson's definition of determinable by extensionality, we cannot ensure the finiteness of the determinable and this leads to an *ad infinitum* return.

In a similar fashion, Searle compares the relation between determinate and determinable to the relation between *genus* and *species*, but again these do not exactly coincide. To explain his point, he uses the notion of predicate entailment, usually understood as a relationship between things that have truth-values. He assumes that determinables characterize things in the same way as determinates do, but that they are only less specific. Suppose, he says, for example the relation red/color. We can replace the determinable 'color' by the adjective 'colored'. In this way, determinables are directly predicable of things. Then, the difference between the two relations is that a *species* is marked off within a *genus* by means of differentia (a conjunctive relation) but in the case of determinate/determinable we have an inclusive relation (the membership to a set). "A species is a conjunction of two logically independent properties –the *genus* and the *differentia*. But a determinate is not a conjunction of its determinable and some other property independent of the determinable. A determinate is, so to speak, an area marked off within a determinable without outside help" (Searle 1959, 143).

Again, this approach seems to be problematic. Though the difference between the distinctions determinate/determinable and *genus*/*species* is different in Johnson and in Searle, the view of statements like 'Plato is a man' as based upon an adjectival predication is unsatisfactory, since there is no adjectival characteristic of an object in possession of a property that is a sufficient and necessary condition for the extensional explanation of the term. In this respect, there would be no difference between the determinate/determinable relation and the class/member relation, with the consequent danger of the unilateral implication that a conjunctive conception of these relationship supposes.

To avoid this accusation, Searle adds a new criterion to his definition. According to this criterion, a determinable specifies a respect of resemblance: "If to say of any two objects that they have the property A entails that they resemble each other (are exactly alike) with respect to B then

A is determinate of B" (Searle 1959, 152). Terms of type A in Searle's account (or determinates) are characteristically adjective-like expressions and terms of type B (or determinables) are abstract nouns. Then, we invent a term (determinable, abstract nouns) to cover a whole range of characterizing terms (determinates, adjectives) which are all "in the same line of business" (Searle 1959, 154). This last characterization fits well with Prior's analysis of determinables as kind of qualities.

According to Prior, terms like 'color' and 'man' are both nouns, not adjectives. The extension of 'color' seems to be the class of colors and the extension of 'man' seems to be the class of men. But whereas men are concrete objects or particulars, colors are properties or universals. Hence color is a property of properties (a second-order property) while a particular instance of color is a first-order property (a primary or essential property). Therefore, 'man' is a first-order term that defines essentially what Plato is. In other words, "when we call redness a color we are saying that it is a particular individual instance of a certain kind of quality. 'Red' is the proper name of an individual universal, if we may speak so; while 'color' is the name of the class of universals to which this individual one belongs" (Prior 1949a, 8). Thus, we can understand a determinable as a name of a class of qualities, or a quality of qualities, that therefore expresses a kind of quality. Thus, determinables refer to respects under which objects are characterized. The ground of collecting several determinates together under a determinable is the special kind of difference (Johnson 1921, 176).

This suggests that determinables can be dimensions of difference or dimensions of comparison, or even quality dimensions (see Hautamäki 1986, 15). According to this characterization of determinables as qualitative dimensions, determinables enable us to express kinds of differences. Consider, for instance, a point in space. The coordinates of this point can be referred to one, two or three axis. Then, we shall say that the point lies on a line, a plane or in the space, respectively. Consequently, we shall have different levels of access to reality depending on the number of dimensions we are neglecting. In other words, the reality outside the set of axis we are using to assign coordinates avoids our analysis.

According to this view, only determinates under the same determinable are comparable. This is to say, Prior defends the relation between determinates and their determinable as something internal to the relationship. When we predicate a certain determinate under a concrete determinable, we are predicating something that cannot be separated from the subject in order to be predicate of it (Prior 1949a, 18). However, Prior also realizes the main difficulty behind this claim: the problem of fitting the relationship into purely conjunctional terms requires justifying the inference from the

conjunction to one of its conjuncts (Prior 1949b, 191-2). Nevertheless, he does not provide an answer to this issue.

The scientific view affirms that the world consists of objects and these objects have properties. This amounts to saying that the world has a determinate/determinable structure. Nevertheless, in the previously mentioned approaches either the relation between determinate and determinable has not been analyzed or its analysis is unsatisfactory. They provide some analysis about the relation between determinates that belong in the same determinable, but establishing this relation is not enough to provide an answer to the ontological status of the determinables and does not answer the question of what a determinable is, as Fine (2011, 184) observes.

As considered by the scientific image, the complete account of such determinate/determinable structure represents the complete information available. In this sense, determinates represent partial information about the object under consideration, a fraction of the total information we would have if we could take into consideration all possible determinables. However, this is not possible since we only have restricted access to determinables according to the particular space of determinates that our point of view is capable of possessing (Hautamäki 1986, 23-8). In this sense, our point of view somehow plays the role of a filter of the physical world. Hence, a particular observer can only have partial, incomplete information about any particular object.

However, this does not mean that this information is something unrealistic. In fact, the scientific image of the world is an accumulation of observations and other objective facts provided by observers, and the particular (subjective) view of these is crucial in establishing this (objective) scientific image. In fact, the particular observers explain the world according to their accessible determinables. These particular observations are about real objects and their aspects (Sellars 1963) and the scientific image exploits this observations looking for the most accurate view of the world. This is in agreement with Hautamäki (1986, 8), where he mentions that even if it is impossible for any particular observer to access the total information of the world, access to a larger number of determinables guarantees a more accurate description of reality.

We shall defend that determinables represent aspects of objects. Then, ultimately two objects are identical or different with respect to their aspects, for instance color. We can select a concrete set of determinates to designate different colors, but we can apply the same set of determinables to two different objects. In this sense, the set of determinables is not exclusively dependent on the object under consideration, but instead on the objective properties of objects in

question. Also it will be impossible to apply to two objects different determinates if they are identical with respect to the quality under consideration. That is, if they resemble one another.

The classical theory of divisions only assumes that the relation between determinable and determinates correspond to a relation between an object and its attributes. This account defines indirectly determinables as classes when claims that two determinates resemble only if they belong in the same determinable. For this reason, we defend that a determinable is a pair consisting of a certain determinable index and a set of determinate values. The determinable index identifies the determinable and represents concepts and the set of the values consists of determinates adequate for the particular index.

Moreover, determinable indices are functors with their associated range of values, and each functor is represented by a function. In this way, a determinable is similar to a variable. The index identifies the domain of definition of the functor and every element of the set of values is an assignable value to the variable, with the particularity that this set of values assigns one and only one outcome to every relevant object according to the range of the index. We also specify a set of values (which we call determinates) of the corresponding function. Then we obtain some flexibility in the application of values according to the range provided by the determinable index. Quoting Carnap, "the classification of attributes into different modalities is not unique but can be made in various ways" (1971, 70-1).

When we evaluate two different determinates, we can define a distance function -or metric, as Carnap (1971, 50-1 and 78-9) calls it- with a numeric value that subsumes the attributes-pair. These two attributes cannot be logically independent; that is, they may belong in the same determinable. Determinates, however, can also be more specific to separate two different objects to which the same determinate is applied. They can represent different shades. The metric for these determinates is then an objective feature of determinables, but totally dependent of the subjective observer's point of view. If this is the case, determinables and determinates under them are universals. In general, following Hautamäki again, we consider determinates as "physical properties even if the determinate systems of natural languages are mental constructs. Determinables are not mental properties, however, because they have an objective status as a part of our language and conceptual schema. Determinables represent objective aspects of real entities" (1986, 23).

In this way, we provide a new approach to the determinate/determinable distinction. Unlike the classical account of divisions, we defend a direct definition of determinables as objective aspects of the objects of the world in contraposition to the definition of determinables solely based on the notion of membership of determinates into the determinable in an extensional way. Hence,

our analysis is not only based on the relation between determinates that belong in the same determinable, but also according to the determination device that permits to establish the relation between the determinable and its possible determinates.

This fact allows to redefine the ontological status of determinables in physical terms. After applying the normal tools that permit to establish the conditions of application from the index set to values, we obtain the relation between objects and observers. These conditions develop a restricted space to relatively determinate states that allow to think about the determinates as values applied to objects according to the concrete conditions or aspects that are relevant to the index under consideration. In contrast to Fine (2011, 184), we suppose that we do not have access to all scenarios. Though he claims the resemblance relation between determinates, his approach seems to lose the specificity required to access to the determinables. That is, given that we only have access to the determinables that appear reflected in our image of the world, we only have access to the determinables after these are determined. Then, the only permitted answer to what this image is will be the set of the total information provided by determinates as physical aspects of objects to which we have access and, by definition, this information is restricted.[1]

REFERENCES

Carnap, R. 1971, 'A Basic System of Inductive Logic', in R. Carnap and R. Jeffrey (eds.), *Studies in Inductive Logic and Probability, vol. 1*, Berkeley, University of California Press, 33-165.

Fine, K. 2011, 'An Abstract Characterization of the Determinate/Determinable Distinction', *Philosophical Perspectives, vol. 25: Metaphysics*, 161-187.

Hautamäki, A. 1986, *Points of View and their Logical Analysis*, Acta Philosophica Fennica Vol. 41.

Johnson, W. 1921, *Logic, part I*, 2nd edition, New York, Dover, 1964.

Johnson, W. 1922, *Logic, part II*, 2nd edition, New York, Dover, 1964.

Prior, A. 1949a, 'Determinables, Determinates and Determinants, I', *Mind* 58:229, 1-20.

Prior, A. 1949b, 'Determinables, Determinates and Determinants, II', *Mind* 58:230, 178-194.

Searle, J. 1959, 'Determinables and the Notion of Resemblance', *Proceedings of the Aristotelian Society*, Suplemmentary Volume 33, 141-158.

Sellars, W. 1963, *Science, Perception and Reality*, London, Routledge.

[1] Project: "Points of View and Temporal Structures" (FII2011-24549).

CAUSAL EXPLANATION IN THEORIES OF PROBABILISTIC CAUSATION

TIM DE CRAECKER
U. of Ghent

ABSTRACT: Ellery Eells introduced the notion of a probability trajectory in the analysis of singular causation. In this paper this notion is further refined in a probabilistic counterfactual framework and related to the work of Jack Good on probabilistic causation. When the search for a causal explanation of an event is motivated by more specific aims, the probabilistic approach is able to select the relevant causes in the causal history of the event, because of its quantitative nature.

Key words: Causal Explanation, Singular Causation, Probabilistic Causality, Counterfactuals.

1 Introduction

Theories of probabilistic causation analyze causality without assuming causal determinism. Unfortunately probabilistic causation is associated with the idea to express causal statements in terms of observed associations or with "the ideal of remaining compatible with the teachings of nondeterministic physics" (Pearl 2000, 249-257). As is well known (Pearl 2000), it is not always possible to distinguish between two sets of causal assumptions on the basis of observed independences alone. For instance, $X \perp\!\!\!\perp Y|M$ is compatible both with M being a common cause, $X \leftarrow M \rightarrow Y$, and with M being a mediator, $X \rightarrow M \rightarrow Y$. To distinguish both empirically one relies on temporal information or on experimental data, describing the effect on M of an "intervention" (a causal notion) on X (or vice versa). In this paper probabilistic causation is motivated along different lines: the use of "probability trajectories" (Eells 1991) in the analysis of singular causation.

1.1 Outline

In section two I introduce the structural account to causation (Pearl 2000) and clarify the distinction between singular and general causal statements. Section three and four contains the main theme of this paper, which is to refine the notion of a probability trajectory. In section three the notion of "objective probability" is analyzed as a stable state of belief (Jaynes 2003). In section four the notion of a probability trajectory (Eells 1991) is further refined and related to the approach of (Good 1961/62). I end this paper with some brief remarks on applications to singular causation.

1.2 Notation

Events will be denoted by lower case letters $x, y, \ldots$, the horizontal bar is used as a symbol for the complement $\bar{x}, \bar{y}, \ldots$. Variables will be denoted by capital letters X,Y,... and are supposed to be binary, $X(\omega) = 1$ meaning "x has happened in ω". The probability distribution on a variable X will be denoted $P(X)$ while $P(X = 1)$ or $P(x)$ denotes the probability that x occurs.

2 Structural Models

2.1 Example: infectious disease

I will give a brief sketch of the structural account of causation (Pearl 2000) using a fictitious example. Imagine a population of N individuals $\Omega = \{\omega_1, \ldots, \omega_N\}$ susceptible to some infectious disease. Suppose we are prepared to apply the following causal law to all individuals in Ω:

> ω would contract the disease (y) if ω eats contaminated food (x), unless ω is immune (u).

Mathematically such causal laws are encoded in structural equations: a causal graph $\mathscr{G}$, together with equations $\mathscr{F}$ relating the endogenous variables to its parents. In this example $\mathscr{G}$: $X \rightarrow Y \leftarrow U$ and $\mathscr{F}$: $Y(X,U) = X \cdot (1 - U)$. The potential outcomes $Y(X,U)$ are summarized in table 1.

Table 1: Potential Outcomes For Y

$Y_{X,U}$	u	$\bar{u}$
x	0	1
$\bar{x}$	0	0

2.2 Singular causation

Singular causal statements are relative to a specific case ω, like "John got ill because he ate contaminated food". Structural accounts of causation use counterfactuals to express such statements. Counterfactuals like "John would get ill if he eats contaminated food" are derived from the description of ω and the structural equations. For instance, $Y_x(\omega)$ is defined as the value Y obtains if X would equal 1, while all non-descendants of Y in $\mathscr{G}$ take the value they obtain in ω.

$$Y_x(\omega) \equiv Y(1, U(\omega)) \tag{1}$$

2.3 General causation

General causal statements are relative to a specific population $\Omega = \{\omega_1, \ldots, \omega_N\}$, like "consumption of contaminated food would increase the prevalence of infectious disease (in Ω)". We assume there is no

interference between the individuals within the population. For instance, in example 2.1 we suppose that it is impossible for individuals $\omega_i \in \Omega$ to infect other individuals $\omega_j \in \Omega$, $i \neq j$. This restriction makes general causation different in nature from singular causation. General causal statements are derived from the structural equations and a probability distribution $P(U)$ on the exogenous variables. This distribution encodes the frequencies with which the exogenous factors occur in the population. Consider again example 2.1, we may want to compare $P(Y_x)$ with $P(Y_{\bar{x}})$. $P(Y_x)$ is defined as:

$$P(Y_x) = \sum_u Y(x,u) \cdot P(u) \tag{2}$$

3 Causal Laws and Objective Probability

3.1 Objective Probability

The probability of a "fair" coin ending up heads is $\frac{1}{2}$ whenever the coin is tossed "vigorously". This first case is one of the most paradigmatic examples of objective probability (or chance). Now, as a second case, imagine we are informed about an experiment, having only two possible outcomes. According to the principle of indifference we have to assign the same belief to both outcomes. In both cases probabilities are the same despite a clear qualitative difference. For instance, only in the first case is it reasonable to place a bet of 5$ on one of the outcomes with a possible return of 10$. This might be a reason to explore alternative approaches to model uncertainty (Halpern 2003). What is precisely the difference between both cases ? In the first case it is not obvious what kind of information would make us change our beliefs. In the second case the information available is very sparse. If anyone were to provide us with information on the experiment this would likely result in a change of our beliefs. Put differently, the belief in the first case is more stable than in the second case. An attempt to model the stability of a state of belief can be found in (Jaynes 2003, 553-588). In roughly the same spirit we can express this stability informally as follows:

> (for the time being) whenever a fair coin is tossed vigorously we will expect heads with belief $\frac{1}{2}$, regardless of anything else we may have been told.

The stability of a belief is never absolute. A detailed analysis of the causal processes involved may still uncover a more subtle system of chances. This is illustrated in the next subsection.

3.2 Case Study: Coin Tossing

It is rooted in habit to use a uniform distribution to model the outcome of a coin toss. The uniform distribution is partly justified by the symmetry between heads and tails. But this justification becomes less compelling if we know the side that was facing upwards when the coin was tossed into the

air. (Diaconis et al. 2007) analyze in detail the process of flipping a coin caught in the hand. The movement of a coin is governed by the law of conservation of angular momentum. The normal of the coin precesses around the angular momentum vector with constant angular velocity ω_N. More precisely, (Diaconis et al. 2007) derive a causal law predicting the outcome of flipping a coin starting heads perfectly up (i.e. initially the normal to the coin is aligned with the up direction). The outcome of the toss is heads if and only if:

$$cos^2(\psi) + sin^2(\psi)cos(\theta) > 0 \quad (\theta := \omega_N \cdot t) \tag{3}$$

In equation (3), ψ denotes the size of the edge between the up direction and the angular momentum and t denotes the time when the coin is caught in the hand (assuming the coin was tossed into the air at $t = 0$). Flipped coins have a tendency to come up the same way they started (Diaconis et al. 2007). This tendency highly depends on ψ. If ψ is $90°$ the coin will be roughly half the time heads up during its flight. However, if ψ is less than $45°$ the coin never turns over.

4 Probabilistic Singular Causation

4.1 Probability Trajectory

The intuïtion behind probabilistic causation is that a cause raises the probability of its effect. As is well known, this idea needs further refinement. Eells (Eells, 1991) emphasized that the idea of "probability change" in singular causation is different than in general causation. In singular causation probability change is understood as a change in the "probability trajectory" of a particular event $Y(\omega)$. Roughly stated, this probability trajectory models how the probability of Y evolved in time. The following examples make this more precise.

4.2 Example: coin tossing

In previous section we discussed the process of flipping coins. The outcome of this process, denoted by $Y(\omega) \in \{1, -1\}$, is determined by three (independent) causes (X, ψ, θ). $X(\omega) \in \{1, -1\}$ denotes whether the coin started heads up or tails up. The causal process evolves in three stages: In the first stage X is determined, afterwards ψ, and θ is determined at the very end of the process.

$$\begin{tikzcd}
\text{Stage 1:} & X \arrow[dr, bend left] & \\
\text{Stage 2:} & \psi \arrow[r] & Y \\
\text{Stage 3:} & \theta \arrow[ur, bend right] &
\end{tikzcd}$$

structural equation:

$$Y = X \cdot sign(cos^2(\psi) + sin^2(\psi)cos(\theta))$$

To model our initial uncertainty we put a uniform distribution on the domain of the variables X, ψ[1] and θ. Before we observe $X(\omega)$, we expect the outcome to be heads with belief $\frac{1}{2}$. However, at each stage we can update this belief. This defines a probability trajectory for Y. As we mentioned in the previous section, flipped coins have a tendency to come up the same way they started. Of course, for some values of ψ it is possible that the coin comes up the other side. If this happens, knowing which side was facing upwards when the coin was tossed into the air actually lowered the probability of the outcome observed.

4.3 Example: infectious disease

Consider again a population of N individuals $\{\omega_1, \ldots, \omega_N\}$ susceptible to some infectious disease. But now the pathogen is transmitted from one individual to another. The ease with which a pathogen is transmitted highly depends on the nature of their relation. For instance, one may more likely contract the disease from a family member than from a stranger. The pathways for transmission of the pathogen are represented in a graph $\mathscr{G}$. For example, in the graph below ω_1 may infect ω_3 through first infecting either ω_2 or ω_4. Let X_i denote whether ω_i is infected or not and let $X_{\mathscr{N}_i}$ denote the values for X of the neighbors of ω_i.

For each pair of neighbors (ω_i, ω_j) in $\mathscr{G}$ we define π_{ij}, the probability that the pathogen is transmitted from ω_i to ω_j if ω_i would get infected. For simplicity, assume that infection can only be contracted from someone within the population. We define π_{ij}:

$$\pi_{ij} := P(x_j(x_i, \bar{x}_{\mathscr{N}_j \setminus x_i})) \tag{4}$$

In a similar way, for each pair of nodes (ω_i, ω_j) in $\mathscr{G}$ we could define $\pi_{ij:p}$, the probability that the infection spreads from ω_i to ω_j through p, a network of pathways, if ω_i would get infected. This seems identical to what (Good 1961/62) called the strength of a network [2] and is a refinement on the notion of susceptibility in epidemiology, i.e. the probability that an individual would contract disease if exposed to the infectious agent. To apply Good's causal calculus we assume there is no interference

[1] (Diaconis et al. 2007) provide an empirical study of ψ
[2] Good's measure of strength seems identical to $-log(1 - \pi_{ij:p})$.

between the links in $\mathscr{G}$, i.e.

$$\forall(\{i,j\} \neq \{k,l\}), X_j(x_i, \bar{x}_{\mathscr{N}_j \setminus x_i}) \perp\!\!\!\perp X_l(x_k, \bar{x}_{\mathscr{N}_l \setminus x_k}) \tag{5}$$

Consider the case where the network is a chain $\omega_1 - \omega_2 - \omega_3$ and ω_1 infected ω_3 through first infecting ω_2. Here, we do observe that x_2 "raised the probability" of x_3, in the sense that $\pi_{23} > \pi_{13}$. This formalizes what (Moore 2009, 121-123) calls the limited transitivity of the causal relation. In (Good 1961/62) it is an axiom.

Stage 1:	$X_1 \xrightarrow{\pi_{12}} X_2 \xrightarrow{\pi_{23}} X_3$	$\pi_{13} = \pi_{12}\pi_{23}$
Stage 2:	$X_2 \xrightarrow{\pi_{23}} X_3$	π_{23}

Consider again the case where ω_1 infected ω_3 through first infecting ω_2, but now ω_1 could also have infected ω_4 although it failed to so in this particular case:

$$\pi_{13} = 1 - (1 - \pi_{12}\pi_{23})(1 - \pi_{14}\pi_{43}) \tag{6}$$

If $\pi_{14}\pi_{34}$ is close to one, while π_{23} is small we have that $\pi_{13} > \pi_{23}$, although x_2 actually caused x_3.

The probability trajectory of $X_3(\omega)$ should be defined as $(\pi_{13:p}, \pi_{23:p})$, with $p : X_1 \to X_2 \to X_3$ the path that is actually causing $X_3(\omega)$. This is in line with (Good 1961/62): each event included in the network p should have a positive tendency to cause its immediate descendants (denoted by a positive sign in the graph above), i.e.:

$$P(\bar{y}(\bar{x})) > P(\bar{y}(x)) \tag{7}$$

This is not the case if we were to include $\bar{x}_4$ in the network p.

4.4 Applications

In the previous examples we knew the causal process and analyzed the corresponding probability trajectory. However, often we would like to do the reverse inference. One method implicit in (Good 1961/62) is to investigate whether there is a chain of events linking the potential cause and the effect such that each event on the chain has a positive tendency to cause the next event. There are at

least two limitations to this approach: firstly, the need to specify the network in sufficient detail and secondly, the presence of interacting causes in the network, like in the example about flipping coins. Another application arises when the search for a causal explanation of an event is motivated by more specific aims and we have to select the relevant causes in the causal history of the event (Halpern et al. 2010). For instance, the doctrine of intervening causation may relieve a defendant from liability, if the harm was caused by an extraordinary natural event or coincidence subsequent to the defendant's act, even though the defendant's act was necessary for the harm (Moore 2009, 229-253). Probability trajectories are useful to clarify these subtleties.

REFERENCES

Diaconis, P., Holmes, S., & Montgomery, R. (2007). Dynamical bias in the coin toss. *SIAM Review*, *49*(2), 211–235.

Eells, E. (1991). *Probabilistic Causality*. Cambridge University Press.

Good, I. (1961/62). A causal calculus i/ii. *British Journal for the Philosophy of Science 11/12*, *305-318*, 43–51.

Halpern, J. (2003). *Reasoning about Uncertainty*. MIT Press.

Halpern, J. & Hitchcock, C. (2010). *Heuristics, Probability and Causality, A Tribute to Judea Pearl*, chapter Actual causation and the art of modeling, (pp. 383–406). College Publications.

Jaynes, E. (2003). *Probability Theory the Logic of Science*. Cambridge University Press.

Moore, M. (2009). *Causation and Responsibility, An Essay in Law, Morals and Metaphysics*. Oxford University Press.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

ACTITUDES COGNITIVAS: EL CASO DE LA ASUNCIÓN, PARTICULARMENTE EN CIENCIA

XAVIER DE DONATO RODRÍGUEZ
Grupo EPISTEME, Universidad de Santiago de Compostela

ABSTRACT: After considering the topic of assumptions as presented by Meinong (1902/1910), we analyze the case of the assumptions in science and propose scientific assumptions as the antecedent of a subjunctive (counterfactual) conditional. This analysis is consistent with Meinong's, where assumptions are not considered to be judgments. For Meinong, "hypothetical judgments" are assumptions rather than judgments (Meinong sees assumptions as something different from judgments). Our analysis also allows to include mere hypotheses as a limiting case, because in this case the conditional would be merely subjunctive (and not strictly counterfactual). Thus, in our analysis, assumptions could be either (i) purely hypothetical, or (ii) counterfactual (in the strictest sense of the antecedent is not satisfied). Thus, the hypothesis would not be considered judgments, but statements of the form "act as if it were true and see the consequences". This sense seems entirely consistent with the normal use of hypotheses in scientific practice.
Keywords: assumptions/suppositions, counterfactuals, hypothetical judgments, idealization, Meinong.

No hay mucha bibliografía sobre la naturaleza y la función de la asunción y otras actitudes cognitivas en ciencia. Bas van Fraassen (1980) y su análisis de la actitud de creer o aceptar una teoría científica y Balzer et al. (1987) desde la perspectiva metateórica del estructuralismo, son ilustres excepciones. De hecho, la asunsción o suposición es en sí mismo un tema muy poco tratado. Alexius Meinong dedicó al tema una de sus obras más importantes (cf. Meinong 1902/1910). Meinong refiere en algún momento a las asunciones en la ciencia, pero que no se detiene a discutir ejemplos ni a proporcionar un verdadero análisis de este tipo de asunciones. En realidad, lo que Meinong pretende se puede describir (él mismo lo hace) como un estudio descriptivo de tipo psicológico de la asunción. No pretende hacer un análisis del concepto ni un estudio lingüístico (aunque en ocasiones lo pueda parecer) acerca de cómo usamos ciertas palabras del lenguaje (tales como "asumamos que…" o "supongamos que…"). Esto es lo que, por el contrario, pretende hacer Roland Hall (1958, 55). Es cierto, que, en el capítulo IV de su libro, Meinong describe varios usos de la asunción (en ámbitos coloquiales y en ámbitos más técnicos), pero su objetivo no es analizar los usos del lenguaje sino los actos mentales aludidos. Tampoco es su intención dar una definición de asunción, por mucho que diga que es un tipo de pensamiento en el que no hay convicción (aunque esto, supuestamente y dada la tipología de experiencias o fenómenos mentales que

Meinong establece, caracteriza unívocamente ese tipo de actos mentales). Mi intención va a ser justamente contraria a la de Meinong e intentar dar con un análisis del concepto de asunción que responda a (la mayoría de) los usos principales tanto en la vida ordinaria como en la actividad científica.

Hablando específicamente de ciencia, resulta muy llamativo que no se haya prestado mucha atención hasta ahora (y que Meinong lo haga en contados pasajes) a la diferencia entre hipótesis y asunciones. Meinong (1902/1910 § 15, p. 82) habla de las suposiciones (*Vermutungen*) y de las hipótesis (*Hypothese*) como de juicios con un grado de convicción más o menos débil, pero enseguida se apresura a aclarar:

> "Es wird dabei freilich nicht immer leicht sein, den Punkt genau namhaft zu machen, wo derjenige, der sich mit einer colchen Hypothese beschäftigt und sie etwa an der Wirklichkeit zu verifizieren unternimmt, aus dem Zustande des Annehmens in den des Vermutens übergeht" (Meinong: *Gesamtausgabe* IV, p. 108).

Interesantemente, en el análisis que daré a continuación, las hipótesis quedan integradas como caso límite de las asunciones. Lo cual, aunque curioso desde una perspectiva como la de Meinong, para quien la asunción es un caso límite del juicio, es una ventaja, puesto que las hipótesis pasan, en mi teoría, a tener en común con la asunción su carácter de no ser juicios (para Meinong las hipótesis, como las suposiciones, sí serían, en cambio, juicios, aunque de un tipo en el que la convicción es muy débil), pasando a diferenciarse de una manera mucho más elemental, que pasaré a explicar enseguida. Aclaremos ya de antemano que esta misma distinción entre hipótesis y asunción es la que parece estar detrás de la distinción de Vaihinger entre suposiciones o asunciones hipotéticas y asunciones ficticias, que habitualmente llama "ficciones". Vaihinger (1911/1924, pp. 144-145) explica que ambos términos expresan dos posibles sentidos del término griego "hypothesis" y que fueron los romanos, por primera vez en la historia del pensamiento, quienes, dándose cuenta de la diferencia, quisieron distinguir explícitamente ambos sentidos con el uso de dos términos diferentes: respectivamente, *suppositio* y *fictio*.

Antes de introducirnos en el tema de la asunción, permítaseme esbozar rápidamente cuál podría ser la diferencia entre presuposición, suposición, supuesto y asunción. En una primera acepción, una presuposición es una suposición implícita en un argumento o una implicatura en el ámbito de un discurso. En este sentido, una presuposición puede ser vista como una premisa implícita en un argumento entimemático. Este el sentido que discute Ennis (1982). Cuando se trata de una implicatura conversacional es lo que Stalnaker (1973) llama "presuposición pragmática". Hay otro sentido aceptable de presuposición, que fue introducido por Strawson (1950 y 1952), según el cual

es un concepto que se refiere a aquellas condiciones que deben satisfacerse para que un enunciado pueda ser considerado verdadero o falso (véase Strawson 1952, 175). Un supuesto es una tesis sustantiva (que puede ser explícita o implícita) que se necesita como premisa en un argumento y que, si el contexto lo aclara, puede no ser necesario explicitar. Una suposición puede verse como una conjetura (aunque no todos los usos de suposición tienen este matiz), una hipótesis para la cual hay muy pocos elementos de juicio. Este es exactamente el sentido que Meinong confiere a la palabra alemana "Vermutung" (que el traductor inglés traduce como "supposition"). En cambio, asunción (*assumption*, *Annahme*) es, en el sentido que Meinong le da, una suposición típicamente *contrafáctica* o para la que no hay elementos de juicio en absoluto; se hace, por así decirlo, en el vacío o, lo que es más habitual, incluso a sabiendas de que lo que se está asumiendo es falso. Esta interpretación puede corroborarse en casi todos los ejemplos y usos que Meinong distingue de la palabra alemana "Annahme". Reconozco que estas distinciones pueden parecer algo arbitrarias, pero son sólo provisionales y para fijar un poco la terminología. De momento, diré que las asunciones meinongianas me parecen en todo punto similares (aunque quizá no idénticas con ellas) a las ficciones de Vaihinger (1911/1924), las cuales son típicamente asunciones de tipo contrafáctico. Este carácter contrafáctico está implícito en el carácter del "como si" que Vaihinger, siguiendo a Kant y su idea de los postulados e ideas regulativas, otorga a las ficciones (*Fiktionen*)[1]. Esto también se constata comparando los ejemplos que Meinong da de asunciones y los que Vaihinger da de ficciones. La terminología está, desde luego, algo enredada, puesto que enmarañados son los usos del lenguaje. Y no sólo me refiero a los términos coloquiales, sino también a la terminología filosófica. Por ejemplo, Walhout (1960) y Rescher (1961) hablan de "suppositions" para referirse a lo que nosotros vamos a llamar, siguiendo a Meinong, asunciones. Mis asunciones corresponden en gran parte, como vamos a ver, a lo que Rescher (1961) llama "suposiciones contravenientes de creencias" (*Belief-Contravening Suppositions*).

Voy con mi análisis, que por razones de tiempo sólo puedo exponer brevemente. Entenderé una asunción científica como el antecedente de un condicional subjuntivo (contrafáctico). En concordancia con lo que dice Meinong, si esto es así, las asunciones no pueden ser juicios, ya que, para Meinong, los antecedentes de los tradicionalmente llamados "juicios hipotéticos" (condicionales, contrafácticos o no) no son juicios en absoluto, sino asunciones (véase Meinong 1902/1910, § 31). Y tiene su sentido que así sea, puesto que, si seguimos a Meinong en su acepción de la asunción como un pensamiento al que le falta en absoluto convicción, ¿cuál puede ser la

[1] Para una reciente exposición del tema de las ficciones en la ciencia, véase Suárez (ed., 2009). Véase también Cartwright (1983).

convicción otorgada a una condición sobre la que no tenemos elementos de juicio o más bien sabemos que no se va a producir o es muy difícil que se produzca?

Las asunciones de tipo contrafáctico corresponden o, más exactamente, dan lugar a lo que la literatura de filosofía de la ciencia ha llamado "idealización" (véase Nowak 1980, McMullin 1985, Laymon 1985, Cartwright 1989, Weisberg 2007, etc.). Conviene distinguir, al igual que en otras actitudes cognitivas[2], la idealización como acto mental u operación mediante la cual asumimos contrafácticamente algo, y la idealización como producto o resultado, que puede ser un enunciado (típicamente, una ley científica), un modelo, una teoría, etc. Tiendo a favorecer una concepción semanticista de las teorías científicas, pero, para simplificar las cosas, hablaré de las idealizaciones como de enunciados o proposiciones que se obtienen como resultado de realizar una serie de asunciones contrafácticas. Más abajo introduciremos cierta terminología al respecto, pero permítasenos antes unas consideraciones previas.

Digamos, para empezar, que las asunciones, tomadas en este sentido, se presentan en una multitud de formas y dan lugar a prácticas establecidas en multitud de dominios y ámbitos del conocimiento. En efecto, lo que Meinong llama "asunciones" y lo que Rescher llama "*belief-contravening suppositions*" pueden adoptar cualquiera de las siguientes modalidades:

- Argumentaciones "por mor del argumento".
- Método de prueba/argumentación por reducción al absurdo: "asumamos que la diagonal de un cuadrado es conmensurable con el lado". "If there were no assumptions, inference would be inexplicable; and this is particularly evident where, as in the case of a *reduction ad absurdum*, premisses and conclusion are actually disbelieved" (Russell 1904, 343).
- Experimentos mentales.
- Situaciones de aprendizaje en las que se incorpora una creencia (Rescher 1961, § III): lo que en otros contextos se llama "revisión de creencias".
- Los tipos ideales de Max Weber contienen usualmente asunciones hipotético-contrafácticas. Así, por ejemplo, *sociedad capitalista* u *homo œconomicus*.
- El uso de simulaciones.

Mi hipótesis, aunque no puedo mostrarlo aquí con detalle ni tampoco ilustrarlo con ejemplos, es que todos estos usos o modalidades de la asunción-idealización tienen una estructura o

[2] Como Ryle (1949, 257) dijo a propósito del juicio como acto mental (lo que Ryle llama "actos judicativos"): "los juicios concluyen con las sentencias; no consisten en sentencias". Evidentemente, nos es legítimo hablar de los juicios como del resultado de ese proceso. Lo mismo en el caso de la asunción y la idealización.

contenido común, que –dada su naturaleza contrafáctica- puede encontrar una paráfrasis formal en la semántica de condicionales contrafácticos de Lewis (1973). El análisis de los condicionales contrafácticos sigue siendo un tema filosóficamente polémico y, en particular, el análisis de Lewis. Yo mismo, en mi tesis doctoral, a la hora de dar un análisis del concepto de idealización, decidí prescindir de los contrafácticos y dar un análisis en términos modelo-teóricos. No obstante, actualmente pienso que, tanto si una semántica lewisiana puede capturar o dar cuenta de nuestros usos de la idealización/asunción como si no, la referencia a condicionales contrafácticos es algo más que un tentador modo de comprender, en lenguaje informal, nuestro uso de las asunciones (idealizaciones).

En general, la idea es que, al imaginar condiciones hipotéticas y contrafácticas, nos estamos alejando más o menos del mundo real o actual. Esta idea se asemeja a la noción de Lewis de la distancia entre los mundos posibles, donde se fija el mundo real por el contexto (por ejemplo, los principios de una determinada teoría o familia de teorías o los conocimientos básicos en un tiempo dado). Lewis dice, por ejemplo, que si hablamos de la necesidad física, nos estamos limitando a mundos en los que las leyes físicas son válidas (los mundos accesibles son aquellos restringidos por las leyes de la naturaleza que se cree satisfechas en el mundo actual). Hablar del "mundo real" no nos compromete en absoluto a ninguna forma de realismo. Lo que, en este contexto, llamamos "mundo real" es un mundo fijado por el contexto y está representado por esa parte de nuestro conocimiento que damos por sentado en un contexto particular (física, biología,... o, más concretamente, si tomamos la biología: la genética de poblaciones, biología evolutiva,...). Supongamos que damos el conocimiento físico por sentado. Con respecto a este mundo, podemos idealizar en diferentes grados: no es lo mismo la idealización de que no hay fricción o que no hay fuerzas externas ejerciendo una influencia relevante en un sistema particular, que la idealización de que los cuerpos son como puntos de masa, donde toda la masa se concentra en un punto infinitesimal que se encuentra en el centro del cuerpo, ya que, según la física clásica, ningún cuerpo ni partícula puede tener una masa igual a cero. En un debate filosófico más amplio, Nozick (2001, 148-155) ha señalado la importancia de distinguir entre diferentes “grados de contingencia”, lo que significa que situaciones que son contingentemente falsas pueden ser verdaderas en mundos posibles que difieren de nuestro mundo real de una manera muy manera diferente (en distinto grado). Voy a utilizar el mismo término, el de grado de contingencia, para subrayar el hecho de que las idealizaciones pueden variar con respecto a su grado de alejamiento del mundo real. La idea de los diferentes grados de idealización ha sido sugerida por los autores anteriores, (cf. p.ej. Jones 2005), pero, ninguno de ellos la ha desarrollado con suficiente detalle.

Este análisis permite incluir como caso límite a las hipótesis, pues los condicionales, en este caso, serían subjuntivos, aunque no estrictamente contrafácticos, en el sentido de que el antecedente no se cumple, ya que en este caso no sabríamos si el antecedente se cumple o no. Así, pues, en mi análisis, las asunciones podrían ser (i) o meramente hipotéticas, o (ii) contrafácticas (en el sentido habitual de no satisfacerse el antecedente). Así, las hipótesis no serían considerados juicios, sino asunciones del tipo: "actuemos como si fuera verdadera y veamos las consecuencias". Este sentido me parece que coincide enteramente con el uso normal que se hace de las hipótesis en la actividad científica, al menos en el momento de ser propuestas, cuando todavía existe apenas ningún elemente de juicio a su favor o en su contra. No solamente eso, lo más interesante de este análisis es que permite explicar el uso explicativo y heurístico de las idealizaciones y, por tanto, de las asunciones hipotético-contrafácticas, pues es el diverso grado de contingencia el que explicaría que dichas idealizaciones puedan en parte encontrar una justificación, pragmática, empírica y teórica: teórica, porque las asunciones más ideales permitirían construir modelos explicativos de distinto grado de idealización; empírica, pues las condiciones menos idealizadas permitirían el diseño de experimentos bajo estricto aislamiento experimental o reducir el campo de aplicación a casos en que dichas condiciones ideales se cumplen aproximadamente; y, finalmente, pragmática, pues las idealizaciones se justifican en la medida en que sirvan para simplificar los cálculos. Esto es justamente lo que autores como Laymon (1985) tienen en mente cuando hablan de la confirmación de idealizaciones. En realidad, de lo que están hablando es de la operación conversa de la idealización, a saber, de la concretización o relajación de condiciones contrafácticas, admitiendo que uno de los factores o parámetros de los que se había prescindido está, de hecho, actuando y ha de ser tenido en cuenta. Un ejemplo muy básico de esto es la ley de Van der Waals y otras restricciones nomológicas en relación con la ley de Boyle-Mariotte en física.

Finalmente y para volver a la asunción y a su conceptualización en Meinong. De acuerdo con el presente análisis, las asunciones (contrafácticas o hipotéticas) no serían propiamente juicios, porque no les puede acompañar el convencimiento de que son verdaderas. Al contrario, en muchos casos (es el caso de las asunciones contrafácticas) se sabe que son falsas o se tienen elementos de juicio para considerar que lo son. En el caso de las asunciones hipotéticas, tampoco, pues no se sabe realmente si son verdaderas o son falsas y, de hecho, a veces parecen implausibles (como sucedió con muchas hipótesis científicas en el momento en que fueron planteadas). Lo que he llamado "idealizaciones" propiamente dichas, e.d. los consecuentes de los condicionales contrafácticos, tampoco son juicios, si los antecedentes no lo son, pues de un enunciado sujeto a condiciones de tipo hipotético-contrafáctico o bien se sabe que no se cumple o que se cumple bajo condiciones que

son muy difíciles de darse. Es justamente en este sentido que Cartwright dijo que las leyes mentían, aunque estrictamente quienes mienten son los científicos que las usan y no las leyes mismas.

Agradecimientos:

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad del gobierno de España (Proyecto de Investigación Ref. FFI2009-08828/FISO).

REFERENCIAS

Balzer, W., Moulines, C.U. and Sneed J.D. 1987, *An Architectonic for Science*, Dordrecht, Reidel.
Cartwright, N. 1983, *How the Laws of Physics Lie*, Oxford, Clarendon Press.
Cartwright, N. 1989, *Nature's capacities and their measurements*, Oxford, Clarendon Press.
Ennis, R.H. 1982, "Identifying Implicit Assumptions", *Synthese*, vol. 51, 1, 61-86.
Hall, R. 1958, "Assuming: One Set of Positing Words", *The Philosophical Review*, vol. 67, 1, 52-75.
Laymon, R. 1985, "Idealizations and the Testing of Theories by Experimentation", en P. Achinstein and O. Hannaway (Eds.): *Experiment and Observation in Modern Science*, (147-173), Boston, MIT Press and Bradford Books.
Lewis, D. 1973, *Counterfactuals*, Cambridge, Mass, Harvard University Press.
McMullin, E. 1985, "Galilean Idealization", *Studies in the History and Philosophy of Science*, 16, 247-273.
Meinong, A. 1902/1910/1977, *Über Annahmen*, en Meinong: *Gesamtausgabe* IV, Graz, Akademische Druck- u. Verlagsanstalt.
Nowak, L. 1980, *The Structure of Idealization*, Dordrecht, Reidel.
Nozick, R. 2001, *Invariances :the structure of the objective world*, Cambridge, Harvard University Press.
Rescher, N. 1961, "Belief-contravening suppositions", *The Philosophical Review*, vol. 70, 2, 176-196.
Russell, B. 1904, "Meinong's Theory of Complexes and Assumptions" (I-III), *Mind*, New Series, Vol. 13, No. 50 (Apr., 1904), 204-219; *part* II: Vol. 13, No. 51 (Jul., 1904), 336-354; *part* III: Vol. 13, 52 (Oct., 1904), 509-524.
Ryle, G. 1949, *The Concept of Mind*, Chicago, University of Chicago.
Stalnaker, R. 1973, "Presuppositions", *Journal of Philosophical Logic*, vol. 2, 4, 447-457.
Strawson, P.F. 1950, "On referring", *Mind* 59, 320-344.
Suárez, M. ed. 2009, *Fictions in Science*, Routledge, London.
Vaihinger, H. 1911/1924, *The Philosophy of "As If"*, trad. ingl. de C.K. Ogden, London, Routledge, 2000.
van Fraassen, B.C. 1980, *The Scientific Image*, Oxford, Clarendon Press.
Walhout, D. 1960, "Suppositions", *The Journal of Philosophy*, vol. 57, 10, 317-326.
Weisberg, M. 2007, "Three kinds of idealization", *Journal of Philosophy*, CIV(12), 639-659.

THE MÜLLER-LYER VISUAL ILLUSION: TRICKING THE BRAIN OR TRICKING THE SCIENTIST?

ALEX DÍAZ
Universidad Autónoma de Madrid

ABSTRACT: Perceptual illusions are generally described as brain misinterpretation of sensory information. This presentation tries to illustrate that, even if the brain misinterpretation hypothesis is only one of the alternatives, it slowly becomes a strong explanation within adjacent fields. It is hypothesized that scientific knowledge is subject to confirmation bias due to the high presence of mainstream-accounts of phenomena. An analytic external account of the Müller-Lyer phenomenon is also enclosed.
KEYWORDS: perceptual illusions, Müller-Lyer illusion, confirmation bias, severe testing, philosophy of perception

Experimental tasks and reflections concerning visual illusions abound in the philosophical and psychological traditions: Aristotle already mentioned the waterfall illusion, Ebbinghaus and Wundt were prolific in their research on perceptual illusions (but see Bates [1923] for a psychological discussion in the early 20th century of the Müller-Lyer display and some historical review of the work on illusion). Even if its nature is not clear, as Bates highlighted, some hypothesis about nature and functioning of illusions are transcending this *hypothesis status* and becoming *strong assumptions*.

Currently, cognitive science is rich in examples where visual illusions (mainly perceptual illusions, such as Ponzo's or Müller-Lyer's) are assumed as brain misinterpretation of sensory information. Paradoxically, this applies to a large domain of fields within psychology, not only perceptual research. A consequence of this *status mutation* is the propagation of some of these assumptions, allowing that a way to understand cognitive phenomena is privileged1.

To illustrate so, a brief compilation is shown: (a) "Cognitive visual illusions refer to perceptual phenomena where sensory information is misinterpreted by the brain" (Renier et al. 2005, 857), in an article concerning illusions and sensory substitution; (b) "visual illusions appear

[1] This tangentially recalls us to Kuhn (1962). The results of derived experimental observations will be more likely to confirm the established theory as they are testing a [new] hypothesis that is somehow contained in the formulation of the experimental task which was, however, supposed to be theoretically unloaded (at least to an extent). This is, that we are building paradigms, not bias-free knowledge.

to require integration at very low levels" (Happé 1996, 876), when discussing that [brain] central coherence is not supposed to be present in children with autistic syndrome disorder and hence they are not affected by illusions; or, (c) "the types of illusions which concern us here are [...] perceptual illusions. They arise from misinterpretation by the brain of sensory information." (Gregory 1968, 279), in an exahustive review and classification of illusions and their implication for the analog-digital distinction in cognitive systems.

According to Gregory, one of the most famous perceptual illusions is the Müller-Lyer figure: "This figure is so simple, and the distortion so compelling that it was immediately accepted as the primary target for theory and experiment" (Gregory 1968, 280). The most-common explanation of the Müller-Lyer illusion is that the angles formed by the arrows are interpreted by our brain as perspective cues (e.g., Gregory, 1968) that compensates usual deformations of 3-D objects on the retinal image. Even if he recognizes that all theorists made developments of their theories with the Müller-Lyer display, which includes Wundt's theory that establishes the distortion as a function of eye movements, not of brain processing; the state-of-art seem to have accepted the *brain hypothesis* as a factual explanation and so did Gregory.

In what follows, I will try to demonstrate, in a fairly analytic way, that Müller-Lyer segments (Fig. 1) are actually different in length. That is, that their length is subject to relativity, which is more related to physics, light properties and retinal structure; than to the subjective computational nature of the brain that is generally assumed.

The argument goes as follows:

1.- There is a limited resolution of the visual system. This is, there exist visual acuity. If we see a pixel and we are not close enough to it, we do not see a black square on a blank ground (i.e., the ideal pixel, Fig. 2), we see a black blurry spot.

2.- This black blurry spot is more similar to a stein, whose structure is supposed to be fractal (Fig. 3). And it is so because the way in which light dissipates in space, and hence reaches our eye, is not completely corpuscular but is subject to perturbations and wave behavior; and also because there exists an inherent constant vibration of the retina. Hence, we can assume that a pixel is never perfectly present in our retina.

3.- If we compare both, an ideal pixel-based retinal image and the more expected fractal-based image (Fig. 4), it becomes difficult to say where the horizontal segment ends in the latter case because adjacent points of light cannot be disentangled from each other. This means that the end of the central segment and the very start of the arrows both become part of the horizontal line.

4.- Then, Müller-Lyer segments are actually different in length when they travel through space and reach a vibrating retina. This is the way in which they are different for our visual system. There is no need that this difference comes from brain processing.

Why, then, the *brain processing hypothesis* has become a *strong assumption* concerning the nature of visual illusions? I just hypothesize that it has been so because of a *theoretical bias.* This may be seen as a form of *confirmation bias*[2] within the scientific domain (see Mynatt, Doherty &

[2] However, this *confirmation bias* usually stands for an individual and content-dependent account of reasoning. I do not want to take this concept in such a strict way but just some of its insight from a philosophy of science perspective. This is, that even if alternative hypotheses are available (e.g., saccadic-system based, sensorimotor accounts, external explanations, or even one concerning the optical properties of the eye), one proper to the mainstream is assumed as explanatory and the rest are ignored when applied on a new field.

Tweny, 1977). This is, cognitive/computationalist accounts of vision became so prolific in the past 50 years that the amount of information, publications, and explanations concerning the nature of visual illusions was simply flooded with brain-processing accounts of visual perception and, hence, they became an assumption when the task was used elsewhere. Then, the computational accounts of illusory effects are privileged and propagate within other fields of knowledge.

The *status mutation* that we mentioned in the beginning of the text (i.e., that a theory or hypothesis suddenly becomes an explanation) has been previously highlighted by a James' student, E. B. Holt in his *The Concept of Consciousness* (1914/1973, 29): "The word *theory* no longer means an hypothesis deliberately added to observed facts but an actually observed law or set of laws from which deductions may be drawn". More recently, more claims arise manifesting that some theories within psychology are becoming explanations (e.g., Leudar & Costall, 2009).

Cognitive science is a wide domain of research, in which hypothesis and explanations easily overlap. This makes it a fertile ground for theoretical biases. When one develops a new experimental task, one might be aware to the extent to which the phenomenon under study has been previously explained or if it just passed some unsevere tests with respect to an hypothesis. Sometimes, it may not be easy to do so because information may flood the publication domain, but we still need to be critic scientists. When applying an old task to a new domain, we scientists are in danger of propagating a wrong explanation of it just because our second-generation hypothesis is much more prone to be confirmed when assuming some of the previous. And, then, the first-generation hypothesis results reinforced when it was not even specifically under observation.

The aforementioned sever testing requirement has been claimed to be crucial for the development of scientific knowledge (Mayo, 2010). Mayo defends the importance of distinguishing areas that are from those that are (thus far) well tested; such a distinction would be blurred if a full theory is accepted when only portions have been well probed. According to her severity principle:

> "Data x (produced by process G) provides a good indication
> or evidence for hypothesis H if and only if x results from a test procedure T which,
> taken as a whole, constitutes H having passed a severe test – that is, a procedure
> which would have, at least with very high probability, uncovered the falsity of, or
> discrepancies from H, and yet no such error is detected".
> (Mayo, 2010, 32)

It is easy for a young science to take previous results as sound and to rapidly run for the next developments. But still, we have to be aware of to what extent previous explanations keep on being tentative hypothesis, which of them cannot be tested at all -because of the nature of both the phenomenon and the explanation-; or if, on the other hand, they can pass some very severe tests, which is expected from any scientific theory. It is true that cognitive phenomena is not always

suitable for the most severe testing desired, but still we have the tools to develop the most ecologically-valid and challenging testing for our hypotheses.

Epilogue: Perception and Illusion

Yet, one might ask: "how do we perceive the end of a line?" Well, several answers are available and paying favor to one or another is not the main aim of this paper. However, we next elucubrate a bit as it relates to the privileged account that was mentioned in the beginning of this paper.

The brain hypothesis comes from a computational account of vision (e.g., Marr & Poggio, 1979). If a computational two-point comparison method (i.e., the one required for a serial computer) is required to infer the end of a line by our visual system, it would be expected that those acute angles in the Müller-Lyer display, would produce no illusory effect -according to the argument we presented-. This is expected to be so because only obtuse angles would make the central segments look longer (i.e., make the arrows an extensive part of the horizontal line).

If, on the other hand, visual systems are sensitive to properties like contrast without inference-processes in a way like Gibson (1979) puts it, and the end of a line is directly specified by contrast (which, let us propose, may be indicated by different and simultaneous inhibition values of adjacent cells); we might hypothesize that acute angles increase the contrast closer to the center of the segment, and that obtuse angles extend the length of the central line and make the highest contrast points be further from the extreme of the central segment. This effect, that acute angles lead to underestimations of length while obtuse angles lead to overestimations, is a robust observation within the Müller-Lyer experimentation (Lewis, 1904; Dewar, 1967; Day & Dickinson, 1976).

However, this is not enough support for a theory, and tons of alternative theories are available when trying to explain so-called perceptual illusions. And there has been observed some additional illusory effects beyond those expected by contrast or size-based models (Restle & Decker, 1977). Some other paradoxical observations include the dissociation between space and size when the experimental task consists of indicating where the central line ends up, which makes the illusory effect disappear. This is, that space judgements about the end of the line are accurate while estimations of length are not (Giliam & Chambers, 1985). Additionally, modifications of the verbal response categories affect the previously observed effect of resistance to illusions in children with autistic syndrome disorder; and this resistance have not been replicated (Ropar & Mitchell, 1999)

Finally, I would like to stress the difficult place that illusions occupy within the experimental research on cognitive science. Probably, they constitute a very challenging testing for hypotheses

while, at the same time, seem not to provide strong support for any theory under observation. I would like to add a remark made by Costall as it points out an important feature of the relationship between illusions and the related scientific knowledge:

> "I think it is important not take the notion of illusion, or the explanations of illusions for granted. For example, what should a stick in water look like?; There's a joke about two people looking down at the street from a tall building. The first says the people down in the street look like ants. The second objects: 'Don't be silly, if they looked like ants from here we wouldn't be able to see them'". (Costall, 2012, personal communication)

REFERENCES

Bates, M. 1923. 'A study of the Müller-Lyer illusion, with special reference to paradoxical movement and the effect of attitude'. *The American Journal of Psychology,* 34, 46-72.

Day, R. H., & Dickinson, R. G. 1976. 'Apparent length of the arms of acute and obtuse angles, and the components of the Muller-Lyer illusion'. *Australian Journal of Psychology, 28,* 137-148.

Dewar, 1967. 'The effect of angle between the oblique lines on the decrement of the Müller-Lyer illusion with extended practice'. *Attention, Perception, & Psychophysics, 2,* 426-428.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception.* Boston, Houghton-Mifflin.

Giliam, B., & Chambers, D. 1985. 'Size and position are incongruous: Measurements on the Müller-Lyer figure'. *Attention, Perception, & Psychophysics, 37,* 549-556.

Gregory, R. L. 1968. 'Perceptual illusions and brain models'. *Proceedings of the Royal Society B,* 171, 279-296.

Happé, F. G. E. 1996. 'Studying weak central coherence at low levels: Children with autism do not succumb to visual illusions. A research note'.

Holt, E. B. 1973. *The Concept of Consciousness*. Arno Press. Reprint of the 1914 edition published by McMillan, NewYork.

Kunh, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press

Leudar, I., & Costall, A. 2009. *Against Theory of Mind*. Houndmills, Palgrave-MacMillan

Lewis, E. O. 1904. 'The effect of practice on the perception of the Muller-Lyer illusion'. *British Journal of Psychology, 2,* 294-306.

Marr, D. & Poggio, T. 1979. 'A computational theory of human stereo vision'. *Proceedings of the Royal Society B, 204,* 301-328.

Mayo, D. 2010. 'Error, Severe Testing, and the Growth of Theoretical Knowledge'. In D. Mayo & A. Spanos (Eds.), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability and the Objectivity and Rationality of Science.* Cambridge, Cambridge University Press, 28-57

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. 1977. 'Confirmation bias in a simulated research environment: An experimental study of scientific inference'. *Quarterly Journal of Experimental Psychology*, 29, 85-95.

Renier,, L. 2005. 'The Ponzo illusion with auditory substitution of vision in sighted and early-blind subjects'. *Perception,* 34, 857-867.

Restle, F., & Decker, J. 1977. 'Size of the Muller-Lyer illusion as a function of its dimensions: Theory and data'. *Attention, Perception, & Psychophysics, 21,* 489-503.

Ropar, D., & Mitchell, P. 1999. 'Are individuals with autism and Asperger's syndrome susceptible to visual illusions?' *Journal of Child Psychology and Psychiatry, 40,* 1283-1293.

DE LOS GÉNEROS CONSTITUIDOS DE KUHN A LOS OBJETOS ABSTRACTOS DE ZALTA*

JOSÉ L. FALGUERA
Universidad de Santiago de Compostela

ABSTRACT: Zalta applies his theory of abstract objects to account for the reference of terms like "(absolute) simultaneity ", "caloric" or "phlogiston", where these references are understood as fictional entities. This conveys that the reference of each of those terms is determined by all the conditions established by the corresponding theory. This proposal is assessed in the light of Kuhn's considerations according to which kinds are the reference of scientific terms.
KEYWORDS: Abstract objects, Zalta, constituted kinds, Kuhn, términos científicos

1. Sobre la Teoría de Objetos Abstractos de Zalta (TOAZ)

Zalta ha desarrollado una teoría formal de objetos abstractos (TOAZ). Con ella defiende, entre otras cosas, que, aunque expresiones como (el número) "2" o "Sherlock Holmes" no denotan entidades existentes espacio-temporalmente (en ningún mundo posible), denotan entidades abstractas. Zalta denomina "objetos abstractos", objetos-A, a entidades que *no pueden ser espacio-temporales* (ya sea porque habitualmente se consideran entidades abstractas, ya porque se conciben como ficticias o como entidades imposibles). Que x es un objeto-A se expresa '$A!x$'. Un conjunto disjunto del de objetos-A en su propuesta es el de "objetos ordinarios", "objetos-O". Que x es un objeto-O se expresa '$O!x$'. Los objetos-O son posiblemente espacio-temporales, o lo que es lo mismo son posiblemente concretos: $O!x=_{df}[\lambda y \Diamond E!y]x$; los objetos-A no pueden ser concretos: $A!x=_{df}[\lambda y \neg\Diamond E!y]x$ (donde '$E!y$' se lee 'y es concreto').

La existencia de esos dos tipos de objetos supone dos maneras de tener propiedades. Los objetos-O *ejemplifican propiedades (o relaciones)*, y esto se expresa formalmente mediante la predicación usual: si a designa un objeto-O y F designa una propiedad que a ejemplifica, esto se expresa mediante 'Fa' (de igual forma '$F^n a_1 ... a_n$' se lee como: $a_1 ... a_n$ ejemplifican la relación 'F^n'). Un objeto-A queda determinado por un grupo de propiedades, con aquellas que lo constituyen. Así, el objeto denotado por "Sherlock Holmes" queda determinado por las propiedades que se le

* Este trabajo participa del proyecto de investigación FFI2009-08828, financiado por el Ministerio de Ciencia e Innovación de España y por el FEDER, del proyecto PICT Redes 2006 Nº 2007, de la Agencia Nacional de Promoción Científica y Tecnológica, de la República Argentina.

atribuyen en las novelas de Conan Doyle (p.e., ser un detective, vivir en Londres, etc.); y un objeto matemático queda determinado por las propiedades que se le atribuyen en determinada teoría matemática. Zalta plantea que un objeto-A *codifica* las propiedades que se le atribuyen (de una forma u otra), y expresa esto formalmente mediante otro tipo de predicación: si '*a*' designa un objeto-A que codifica la propiedad designada por '*F*', eso se expresa mediante '*aF*'. Pero mientras los objetos-O sólo ejemplifican propiedades, los objetos-A además de codificar, también ejemplifican propiedades. P.e., Sherlock Holmes ejemplifica las propiedades de 'ser personaje de ficción', 'ser concebido por Conan Doyle', etc.; un objeto matemático ejemplifica las propiedades de 'ser abstracto', y de 'no tener masa', etc.

La idea es considerar como objetos-A a entidades sobre las que hay alguna representación (por ejemplo, mediante un relato de ficción, mítico, de teoría matemática), debido a que son el objeto intencional de esa representación, y que no es posible que sean espacio-temporales. En todo caso, aunque los objetos-A hasta aquí considerados han sido individuos, Zalta también contempla que ciertas propiedades y relaciones sean objetos-A: x^t es abstracto$^{\langle t\rangle}$ $[A!^{\langle t\rangle}\ x^t] =_{df} [\lambda y^t \neg\Diamond E!^{\langle t\rangle}\ y]x^t$ (siendo $\langle t\rangle$ el tipo apropiado para propiedades que tienen como instancias objetos de tipo t). Así, las propiedades designadas por expresiones como "unicornio" o "hobbit", dado su carácter ficticio, son objetos-A (según TOAZ). Lo mismo sucedería con términos científicos como "flogisto" o "calórico". Además, podemos aceptar que tales propiedades son universales con sus instancias, y éstas a su vez son ficticias, y por ello a su vez objetos-A (pensemos, por ejemplo, que en cierto relato de ficción cabe contemplar determinadas instancias particulares de hobbit, como "Bilbo" y "Frodo").

Zalta (1983, VI-3.B) proporciona un ejemplo interesante de relación que, en tanto que ficticia, debería considerarse abstracta; a saber, la correspondiente al término "simultaneidad (absoluta)". Es ficticia a la luz de que queda rechazada por la Teoría de la Relatividad. Contemplar que "simultaneidad (absoluta)" denota una entidad ficticia permite a Zalta proporcionar un análisis ontosemántico en el que cobre sentido hablar de "la no existencia de la simultaneidad" (en concreto, para un análisis de la oración "Einstein descubrió que la simultaneidad no existe", toda vez que la no-existencia *espacio-temporal* es ejemplificada en este caso por el objeto-A que designa la expresión "simultaneidad") (cf. Zalta 1983, 146). Siguiendo esa consideración, también podemos decir, por ejemplo, que relaciones funcionales, como la correspondiente a "masa newtoniana", son objetos-A.

Los objetos-A (sean particulares, propiedades o relaciones) son incorporados en TOAZ mediante un *esquema de axioma de comprensión,* que establece que para una condición dada φ

sobre propiedades de cierto tipo $\langle t \rangle$ hay un objeto-A que codifica precisamente tales propiedades y no otras: $\exists x^t(A!^{\langle t \rangle}x \& \forall F^{\langle t \rangle}(xF \equiv \varphi))$ [donde φ es cualquier fórmula en la que x^t no está libre]. Además, para cada condición ese objeto-A es único, según un teorema (de unicidad) de TOAZ

2. Lo que denotan los términos científicos conforme a TOAZ

Los tratamientos dados a las propiedades designadas por "flogisto", "calórico", "simultaneidad", "masa newtoniana", etc. a la luz de TOAZ serían semejantes a los que desde dicha teoría cabría dar a entidades ficticias como las denotadas por "Sherlock Holmes" o "Sancho Panza". Las últimas son entidades individuales constituidas por los respectivos relatos (novelas) en las que se caracterizan los personajes correspondientes a dichos nombres propios, las primeras son propiedades o relaciones abstractas constituidas por las respectivas teorías en las que se caracterizan las entidades, ahora consideradas ficticias, correspondientes a las diferentes expresiones científicas.

Pero cierto punto de vista acerca de las teorías empíricas propicia ampliar ese tratamiento a cualesquiera términos científicos. Basta adoptar cierto anti-realismo epistémico acerca de las teorías empíricas y sus términos científicos para poder plantearse conforme a TOAZ que los términos científicos en general denotan objetos-A (propiedades y relaciones, en este caso)

Tres consideraciones son pertinentes al respecto:

1. Toda teoría empírica está supeditada a ser finalmente rechazada, y con ello a considerarla como pudiendo proporcionar una manera incorrecta de dar cuenta de cómo o son ciertas parcelas del mundo. Es más, en tanto no haya una buena justificación para considerar correcta una teoría, puede considerarse como una ficción útil en su parte teórica. Si así fuera, estaría justificado considerarla como un relato ficticio y aplicarle el anterior tratamiento a lo que denotan tales expresiones.
2. En el mejor de los casos, las teorías se consideran como proporcionando idealizaciones o propuestas aproximadas acerca de parcelas del mundo. Cuando así es, las entidades designadas por sus expresiones descriptivas teóricas proporcionan propiedades (o relaciones) inexactas, no reales. Por ello tales propiedades (o relaciones) son objetos-A, con la peculiaridad de que a tales propiedades (o relaciones) pueden corresponderles respectivos objetos-O (del mismo tipo). (Pero tal correspondencia no entrañaría que es el mismo objeto, ya que un objeto o bien es abstracto o bien es ordinario.)
3. Lo dicho anteriormente en (2) es aplicable a los términos descriptivos *T*-no-teóricos de cada teoría empírica *T*, si cada uno de ellos es teórico para alguna otra teoría.

Conforme a esta interpretación cabe pensar en las teorías empíricas a semejanza de los relatos de ficción que con sus diferentes términos descriptivos dan cuenta de objetos-A, al modo en que Zalta lo hace con la Mecánica Newtoniana para analizar la relación designada por "simultaneidad (absoluta)". De ser así, TOAZ no sólo sería aplicable a términos como 'flogisto', 'calórico', etc. sino a cualquier término descriptivo de la ciencia empírica.

Sin embargo, la aplicación de TOAZ a términos científicos, para establecer sus respectivas referencias, no está exenta de problemas. Para constatarlo, atendamos a de una propuesta que quepa catalogar de anti-realista y que *a fortiori* contemple que cualquier término científico tiene como referencia una entidad ficticia (y, por ello, abstracta), como la desarrollada por Kuhn en sus últimos trabajos de filosofía de la ciencia.

3. Los géneros constituidos de Kuhn

Kuhn, en el marco de proporcionar una elucidación de la semántica de los términos científicos que diera cuenta de los casos de inconmensurabilidad interteórica –entendida como variación conceptual entre los respectivos términos de ciertas teorías con divergencia de referencia–, planteó que la referencia de los términos científicos viene dada por géneros [*kinds*]. Los géneros para Kuhn son constituidos por la acción humana (cf. Kuhn 1991, 4; y 1993, 315) –no son 'géneros naturales', en la acepción realista de esta expresión–, aunque a veces *pudieran corresponder* a maneras en que realmente está dividida la naturaleza. En tanto que constituidos, y dada la posición anti-realista de Kuhn, podríamos decir que los géneros considerados por Kuhn *no* son objetos-O sino objetos-A (conforme a las definiciones de Zalta). A fin de evitar equívocos a partir de ahora llamaré "géneros constituidos" a los géneros de los que habla Kuhn. Por otro lado, debe tenerse en cuenta que para Kuhn entre los términos de género no sólo hay términos clasificatorios, sino también de términos comparativos o métricos (entre los últimos como ejemplos paradigmáticos están los términos de la mecánica clásica "fuerza", "masa", "peso" (cf. Kuhn 1993, 317).

Hacking (1993), interpretando a Kuhn, ha propuesto entender los géneros que éste toma en consideración al modo nominalista, identificándolos como conjuntos, es decir, entidades extensionales. Pero esa interpretación tropieza con una seria dificultad. Resulta que cada una de dichas entidades extensionales para un término científico es determinada por el conjunto de leyes (y métodos de determinación para ese término) que conforman la teoría empírica para la que dicho término es teórico. Pero el conjunto de leyes (y métodos de determinación) de una teoría compleja suele cambiar con el desarrollo de la teoría, porque con tal desarrollo se incorporan nuevas leyes o de rechazan algunas previamente asumidas. Si las referencias de los términos científicos son sus

extensiones, ello conllevaría que la referencia de cada término científico podría cambiar con el desarrollo de una teoría (y no meramente al cambiar de teoría por otra inconmensurable). Tal consecuencia no es compatible con el tipo de semántica que perseguía Kuhn, que asumía que no hay variación ni de significado ni de referencia de un término científico en tanto se use en el marco de una misma teoría y a través de su desarrollo. Así pues, si para Kuhn las referencias de términos científicos las proporcionan géneros, éstos no deben ser entidades extensionales que cambian con el desarrollo de las teorías.

Más prometedor, para entender las propuestas de Kuhn, es el criterio adoptado por Bird (2000, V), conforme al que habría que diferenciar entre la referencia y la extensión de los términos científicos.[1] Al respecto hemos de tener presente que, dada una teoría *T*, si bien la extensión efectiva de un término que es *T*-teórico viene determinada por el conjunto de leyes (y métodos de determinación para ese término) relativamente a cada momento de desarrollo de *T*; sin embargo, el género constituido, que es la referencia de dicho término a lo largo del desarrollo de *T*, ya resulta determinado con ciertas condiciones fundamentales, en tanto que se consideran en relación a las aplicaciones paradigmáticas de *T*. Nótese que tales condiciones y aplicaciones se preservan a través de todo el desarrollo de *T*. Kuhn presentó esas condiciones fundamentales como generalizaciones con el carácter de *sintético a priori no absoluto*. En Falguera (2011) he defendido que corresponde tal carácter a la(s) ley(es) fundamental(es) –en el sentido de la metateoría estructuralista (cf. Balzer *et al.* 1987)– de una teoría compleja (la segunda ley de la mecánica clásica es un caso paradigmático de ley fundamental).

En todo caso, atendiendo a la mencionada distinción podemos pensar que la referencia de un término científico es un género constituido, es decir, un objeto abstracto (no-extensional) que se caracteriza por quedar determinado mediante cierta condición. Esa condición (que especifico más adelante) es establecida mediante la teoría para la que el término es teórico (conforme a una distinción teórico/no-teórico relativizada a cada teoría).

La idea de que los géneros constituidos se establecen a la luz de cierta condición, parece aproximarnos a TOAZ. Podríamos pensar que la condición en cuestión nos permite aplicar el esquema de comprensión de TOAZ para determinar el objeto-A que según Kuhn denota cada término científico. Pero tal proceder sería incorrecto.

[1] Aunque Bird considera tal distinción a fin de cuestionar la solvencia de la semántica que atribuye a Kuhn , crítica ésta que no vamos a considerar aquí (véase Falguera 2004).

4. TOAZ y géneros constituidos

Consideremos cualquier teoría empírica T como un relato, como en TOAZ con la Mecánica Newtoniana. Tendríamos que los géneros constituidos denotados por términos T-teóricos quedarían determinados por *todo* lo que se les atribuye según T; de igual forma que los objetos-A denotados por "Bilbo" o "hobbit" quedan determinados por las propiedades que cada una de tales entidades codifican, y tales propiedades son *todas* las que se les atribuye *según el relato (o relatos) en que se habla de* "Bilbo", o de "hobbit". Así, sea Σ_T un operador que se lee "según (en) el relato (teoría) T", y mediante el que podemos establecer las entidades tal y como son concebidas internamente a T (como en TOAZ). En tal caso, dada una teoría T, donde $r_1,\ldots,r_m$ son sus términos T-no-teóricos y $s_1,\ldots,s_n$ son sus términos T-teóricos, conforme a TOAZ tendremos que: si $s_1^T,\ldots,s_n^T$ nombran los géneros constituidos denotados por los términos T-teóricos, entonces, haciendo uso del axioma de comprensión de objetos-A y del teorema de unicidad, cada $s_i^T = \iota z \forall G(zG \equiv \Sigma_T G s_i)$.

Pero la asunción del axioma de comprensión es demasiado fuerte para captar la intuición de Kuhn, dado que establece que el objeto-A denotado por un término T-teórico es exactamente el que codifica *todas* las propiedades que se le atribuyen según T. Y esto exige considerar todas las condiciones que conformen T (supuesto que el término en cuestión figura en ellas).

Sin embargo, según la anterior interpretación del punto de vista de Kuhn, tenemos que para determinar los géneros constituidos denotados por los términos T-teóricos no es preciso considerar todas las condiciones que conformen T (a lo largo de su desarrollo). Recordemos que los géneros constituidos ya quedarían determinados con la(s) ley(es) fundamental(es) de T [donde FL_T expresa tal(es) ley(es) fundamental(es)], junto con la formulación lingüística de los datos T-no-teóricos de las aplicaciones paradigmáticas de T [donde, siendo I_p las aplicaciones paradigmáticas de T, $\mathcal{A}(I_p)$ está por la formulación lingüística de tales datos]. Es decir, los géneros constituidos denotados por los términos T-teóricos quedarían determinados con $FL_T \& \mathcal{A}(I_p)$.

Digámoslo formalmente. Sea $FL_T = FL(r_1,\ldots,r_m, s_1,\ldots,s_n)$, entonces el resultado de reemplazar en $FL_T \& \mathcal{A}(I_p)$ un término T-teórico s_i por una variable y_i, tal que y_i quede ligada por el operador λ, nos proporciona la condición que determina cada s_i^T $(1 \leq i \leq n)$.

Esto es:

para s_1^T esa condición es $[\lambda y_1(FL(r_1,\ldots,r_m,y_1,s_2,\ldots,s_n) \& \mathcal{A}(I_p))]$

para s_2^T esa condición es $[\lambda y_2(FL(r_1,\ldots,r_m,s_1,y_2,\ldots,s_n) \& \mathcal{A}(I_p))]$

…

para s_n^T esa condición es $[\lambda y_n(FL(r_1,\ldots,r_m,s_1,s_2,\ldots,y_n)\&\mathcal{A}(I_p))]$

Con lo que, si para s_i^T representamos la condición por G_i, tenemos que $G_i=[\lambda y_i(FL(r_1,\ldots,r_m,\ldots,y_i,\ldots)\&\mathcal{A}(I_p))]$. Y entonces $s_i^T=\iota z\ (zG_i\ \&\ (\Sigma_T G_i s_i)\quad [1\leq i\leq n]$.

Así, cada s_i^T $(1\leq i\leq n)$ está nombrando la propiedad o relación abstracta correspondiente al i-ésimo término *T*-teórico que se relaciona al resto de los objetos-A correspondientes al resto de los términos (*T*-teóricos y *T*-no-teóricos) según la(s) ley(es) fundamental(es) de *T*, cuando ellas son consideradas para las aplicaciones paradigmáticas de *T*.

Sin embargo, según TOAZ esto no es aceptable. A la luz de TOAZ, si s_i es *T*-teórico su referencia se determina mediante *todas* las propiedades que se atribuyen en *T* a lo largo de su desarrollo, y no mediante una propiedad nuclear como sería "$\lambda y_i(FL(r_1,\ldots,r_m,\ldots,y_i,\ldots)\&\mathcal{A}(I_p))$". Pero, ¿significa eso que un término *T*-teórico carece de referencia en tanto no se ha completado el desarrollo de *T*? Aceptar tal cosa supondría asumir, por ejemplo, que en una teoría como la mecánica clásica términos como 'masa' o 'fuerza' carecían de referencia cuando Newton publicó "Philosophiae Naturalis Principia Mathematica" porque nuevas leyes fueron añadidas posteriormente, como la de Coulomb. Eso resulta poco satisfatorio.

Una alternativa supondría considerar que con cada nuevo desarrollo de una teoría, como la Mecánica Clásica, sus términos teóricos tienen nuevas referencias. Pero tal planteamiento tampoco resulta satisfactorio, ya que su coste es no diferenciar entre el cambio de una teoría a otra (como el de la Mecánica Clásica a la Teoría de la Relatividad) y cada desarrollo de una misma teoría (como los de la Mecánica Clásica). El coste metateórico no es asumible. Téngase presente que bajo este enfoque, la mera preservación de alguna(s) generalización(es) fundamental(es), cuando se añaden nuevas generalizaciones especiales (por ejemplo para nuevas aplicaciones), no parece suficiente para decir que hay continuidad de teoría, si se asume que sus términos no preservan las referencias.

Finalmente, otra alternativa es aceptar que cada término *T*-teórico denota un objeto-A y que éste es fijado por una propiedad nuclear –la que se expresa mediante "$\lambda y_i(FL(r_1,\ldots,r_m,\ldots,y_i,\ldots)\&\mathcal{A}(Ip))$"–, aunque ese objeto pueda codificar otras propiedades. Las dificultades que entrañan las otras opciones mencionadas hacen que esta última adquiera plausibilidad. Pero si es así, tenemos que cabe considerar que la TOAZ no es apropiada para tales objetos-A.

Conclusión

Si, conforme a las propuestas de Kuhn, dada una teoría *T*, los géneros constituidos –en tanto que objetos-A– son aceptados como referencias permanentes de los términos *T*-teóricos, tenemos que TOAZ no da cuenta de ellos. Ni siquiera de los términos *T*-teóricos de una teoría considerada inadecuada.

REFERENCIAS

Balzer, W., Moulines, C.U., Sneed, J. 1987, *An Architectonic for Science. The Structuralist Program,* Dordrecht, Reidel.

Bird, A. 2000, *Thomas Kuhn,* Chesham Acumen Pub. Lted..

Falguera, J. L. 2004, "Las Revoluciones Científicas y el Problema de la Inconmensurabilidad", en: González, W.J. *Análisis de Thomas Kuhn: Las revoluciones científicas,* Madrid, Trotta, 177-223.

Falguera, J. L. 2011, "Consideraciones de índole ontoepistemosemántica", en *Metatheori**a***, 1 (2), 39-62.

Hacking, I. 1993, "Working in a New World: The Taxonomic Solution", in Horwick, P. (ed.) 1993, *World Changes. Thomas Kuhn and the Nature of Science,* Cambridge (Massachusetts), The MIT Press, 275-310.

Kuhn, T. S. 1991, "The Road Since Structure", in: Fine, A., Forbes, M., Wessels, L. (eds.) 1991, *PSA 1990, 2 Vol.,* Michigan, East Leasing, 2-13.

Kuhn, T. S. 1993, "Afterwords", in Horwich, P. (ed.) 1993, *World Changes. Thomas Kuhn and the Nature of Science,* Cambridge (Massachusetts), The MIT Press, 311-341.

Zalta, E. N. 1983, *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Dordrecht, D. Reidel

Zalta, E. N. 1988, *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, MA and London, The MIT Press.

READING "DE INTERPRETATIONE" 9

MARÍA JESUS GARCÍA-ENCINAS
Universidad de Granada

ABSTRACT: I will offer here a non-conventional reading of *De Interpretatione* 9. I will use the Aristotelian distinction between potency and act and his understanding of powers as potencies to favor my reading. There are different proposals in the literature about which is in fact the view that Aristotle is finally defending in this text. I want to propose what I take to be a different reading, along these lines.

The future that is contingent exists in potency. It is contingent because predicted things are, in potency, as predicted or as predicted to be not. So statements about future contingents are potentially true and potentially false. That statements about the contingent future are potentially true means that the things they predict can be in the future as they predict; that they are potentially false means that the things they predict can be in the future contrary to prediction. How things will actually be depends on the exercise of powers/potencies. And these can be either activated or not activated in the future.
KEYWORDS: Aristotle; future contingents; potency; truth

1. Being potentially true and potentially false in Metaphysics V

In *Metaphysics* V, 7 Aristotle claims that being potentially and being actually apply to the three divisions of being: (1) being *per accidents*, that is, in the accidental way, as when a man is musical; (2) being *per se*, that is, when things are in their own right, as when a man is a human being; (3) and being true/false, that is, when statements or their negations are true or false. ($1017^{a}35$)[1]

Aristotle then gives some examples of the distinction between potency and act applied to the accidental being and being *per se*: something sees, both, potentially and in act; something knows both, potentially and in act; and something rests, both, potentially and in act ($1017^{b}1$-5). And it seems clear, given his immediate affirmation before, that he accepts that statements are true, both, potentially and in act; and that statements are false, both, potentially and in act.

2. Being potentially in Metaphysics IX

In *Metaphysics IX* Aristotle explores the distinction between being potentially X and being

[1] This is also noted by Charlotte Witt (1989, 129).

actually X, and he makes a distinction between agent powers and patient powers. He says that substances and objects have the potential to do things other than they are doing at present, or to be other than they are at present.

So there are *agent powers* or capacities like a person's power to walk or see, the house builder's art of house building, fire's power to burn,…. And there are *passive powers* or capacities, like the power of an olive to be crushed, the power of a patient to be cured, the power of a white thing to be black, … When appropriate conditions are in place, powers are activated or expressed. Absent the appropriate conditions, powers are inactive. So powers can be *active* or *dormant*. A dormant power is a being that exits potentially it is to account for the present existence of inactive powers that Aristotle will use the concept of potentiality (Witt 2008, 13). When an agent power like walking is inactive it exists potentially; and when it is active, it exists actually. And powers are directed towards actuality.

I want to apply these ideas about *passive* powers that are in a *dormant* state, or inactive, to the being potentially truth/false of a statement. And to do this I need to mention some non-controversial ideas about Aristotelian truth.

3. Aristotle on Truth: some notes

*Truth i*s saying of what is the case that it is the case, and of what is not the case that it is not the case; falsity is saying of what is the case that it is not the case, and of what is not the case that it is the case. (*Metaphysics IV,* 10, 1011^{b}25-28) Truth makers are things in the world; truth-bearers are statements (affirmations and negations; not sentences). An affirmation is true if the appropriate entities are combined: "Socrates is white" is true iff Socrates and Whiteness are combined. A negation is true if the appropriate entities are *NOT* combined: "Horses are not rabbits" is true iff Horse and Rabbit are divided; contrariwise for false statements. (*Metaphysics E,* 4, 1027^{B}20-24.)

Also, there are elements that can sometimes be combined and sometimes be divided: sometimes Socrates is walking; sometimes he is not walking. Aristotle maintains that some statements can be true at a time and false at another. "Socrates is walking" true at noon, but false early in the morning. Statements change in truth-value because their truth-values are a *purely relational feature*; *entirely fixed by the states of the world*. If the world has contrary properties at different times, then the statement varies in truth-value accordingly. So statements can change in truth-value, even if they do not naturally change, like substances change. (*Categories* 4^{a}21-4^{b}5). That is, their change in truth-value does not affect their nature. Statements are then not agents for their truth-values; so they have the power to change in truth-value, but this is a passive power.

Moreover, Aristotle claims in different parts of his work that some statements are contingently true/false: they are possibly true when they are false; and they are possibly false when they are true. Consider, for instance, Socrates power to walk. His power is activated when he is walking, and then "Socrates walks" is true. But it is true contingently, as Socrates could be not walking. So we can say that "Socrates walks" is contingently true, and also true in act. And when Socrates's power to walk is not activated, and Socrates does not walk, "Socrates walks" is false in act, and contingently false.

If we apply all these ideas to statements about future contingents, I think that we get that "Socrates will walk" is potentially true given that Socrates has the power to walk and that he can exercise it. But also, given that it is also possible that he remains at rest and does not exercise his power to walk, it is potentially false that "Socrates will walk". So that "Socrates will walk" has its truth-value only in potency. For the future that it predicts does not exist in act. "Socrates will walk" is not actually true/false, but it is potentially true and potentially false. Moreover, as Aristotle thinks that being in potency is being directed toward actuality, potential truth and potential falsity come in degrees.

Let's now try to apply this reading to future contingents in *De Interpretatione 9*.

4. Reading De Interpretatione with the aid of powers

There are, I think, two main theses that Aristotle is holding at the same time in this text. One of them is the idea that if *Tertio Excluso* applies to statements about the future, then, it cannot be contingent, but occurs of necessity. So, if for any statement it is necessary that the affirmation or the negation is true, and statements about the future are always true, then they are necessarily true; for their truth-value is unchangeable. That is, if a statement is *always* true or false, then it is necessary (given *Tertio Excluso*). And I think that these two passages confirm this Aristotelian belief:

> "If it is white now it was true to say earlier that it would be white; so that it was *always true* to say of anything that has happened that it would be so." (18b10) "But if it was *always* true to say that it was so, or would be so, it could not be not so, or not be going to be so." (18b9)

And also:

> "If in *the whole of time* the state of things was such that one or the other was true, it was necessary for this to happen, and for the state of things always to be such that everything that happens, happens of necessity. (19a) "For *what anyone has truly said would be the case cannot not happen;* and of what happens it was *always true* to say that it would be the case." (19a5)

So if a statement is always true, then, given *Tertio Excluso*, its truth-value cannot be changed and then, it is true of necessity. I do not want to question this Aristotelian reasoning. As my goal is to see his solution to the puzzle, let's just accept this ideas for the sake of the argument.

There is also another important thesis that Aristotle defends in the text: that all statements, including statements about the future, have some truth-value. They must have some truth-value, because, if they did, nothing would be one way or another. And there two passages where this idea appears. One is this:

> "If it is not the case it is false, if it is false it is not the case. So it is necessary for the affirmation or the negation to be true. It follows that nothing either is or is happening, or will be or will not be, by chance. (…) it might equally well happen or not happen, since what is as chance has it, is no more thus than not thus, *nor will it be*." (18b5)

So if statements about the future do not have a truth-value, then nothing would be (or not be). A clearer statement of this, is this other passage:

> "Nor, however, can we say that neither is true – that it neither will be nor will not be so. For, firstly, though the affirmation is false, the negation is not true, and though the negation is false, the affirmation, on this view, is not true." (18b17-18b18)

That is, suppose that neither the negation nor the affirmation is true. Then, if the affirmation is false, the negation cannot be true (by supposition). And again, if the negation is false, then the affirmation cannot be true (by supposition). So if statements about the future are not true, nothing will be.

So we have two main theses that Aristotle wants to conciliate in the *De Interpretation*. Now, let's force the distinction between potential being and being in act into the text. There are some passages that can be read under this light. For instance, when he says that there are things that happen contingently:

> "For we see that what will be has an origin both in deliberation and in action, and that, in general, in things that are not *always actual* there is the possibility of being and not being; here both possibilities are open, both being and not being, and consequently, both coming to be and not coming to be." (19a7)

There are things that happen because we do them through deliberation. That is, we have the *(active) power* to do them. And there are things that are not always in act. So they exist some times *in potency*. And for these, there is the possibility of being and not being. These things can be; and can not-be. So powers/potencies explain and allow for both. Or consider this other passage --where he considers the possibility that a cloak will be either cut up or wear out:

> "It is possible for this cloak to be cut up, and yet it will not be cut up but will wear

> out first. But equally, its not being cut up is also possible, for it would not be the case that it wore out first unless its not being cut up were possible." (19a12)

This is the idea: The cloak can be cut up. So it is possible that it is cut up. This is compatible with the supposition that the cloak will wear out first. But then, if it wears out, it is also possible that is not cut up. So both possibilities, that it is cut up and that it is not, are open.

Now let's introduce powers. We can say that the cloak has the passive power to be cut up; and this explains why it can be cut up or be not cut up in the future. So that something is possible (even if it never happens) is explained by its being in potency. And for that that exits in potency, both contraries are allowed (that it will be cut up and that it will not be cup up).

Can we also introduce the *being potentially true* of the statements about the future of the cloak? I think that we can. And the final Aristotelian solution to puzzle about the contingency of the future, is to maintain that both contraries can happen because the future exists in potency –and this allows its contrary to be also possibly the case in the future. So that both contrary statements about the contingent future are both true in potency, but never in act. Let's then try this reading.

> "For contradictories: everything necessarily is or is not, and will be or will not be; but one cannot divide and say that one or the other is necessary. I mean, for example: it is necessary for there to be or not be a sea battle tomorrow; but is not necessary for a sea-battle to take place tomorrow, nor for one not to take place – though it is necessary for one to take place or not to take place." (19127)

This is my proposal: *Tercio Excluso* applies, of necessity, to whatever is in act. But this is not to say that whatever is in act is necessarily /always in act, for before tomorrow, the sea-battle exists only in potency. Tomorrow, necessarily, there is a sea-battle or there is none. So it is *always true that there is* a sea-battle (when there is a sea-battle), and it is always false that there is not (when there is not). But even if there is a sea-battle tomorrow, it is still the case that the now being in potency of the see-battle allows for the possibility that there is no sea-battle tomorrow. So *Tertio Excluso* does not apply to what is in potency. And then, *it is potentially true that there will* be a sea-battle tomorrow; and it is true in potency that there will not be a sea-battle tomorrow. And Aristotle has already written that:

> "Not everything is or happens of necessity: some things happen as chance has it, and of the affirmation and the negation neither is true rather than the other; with other things it is one rather than the other and as a rule, but still it is possible for the other to happen instead." (19a 9)

Concerning chance, neither contrary is more true rather than the other, because in real

chance both can equally come to be. So both, the affirmation and the negation, in future terms, are equally true. But in some cases, as in cases where there are rules of nature, one of the contraries is more possible than the other; so one of them is more true than the other (even if ti is still possible that the other will happen). Potential truth/falsity (like potency) is actualized by degrees.

> "… with things that *are not always* so or are *not always* not so. With these it is necessary for one or the other of the contradictories to be true or false – not, however, this one or that one, but as chance has it; or for one to be true rather than the other, *yet not already* true or false." (19a32)

It is not already true (or false) that there is a sea-battle; rather, *tomorrow will be true in act* that there is a sea-battle (if there is a sea-battle); and it will be *true in act* that there is not (if there is not). But as now the sea-battle only exits in potency, it is true in potency that there will be a sea-battle tomorrow; as it is true in potency that there will be not a sea-battle tomorrow.

> "Clearly, then, it is not necessary that of every affirmation and opposite negation one *should be* true and the other false. For what holds for things that are does not hold for things that are not but may possibly be or not be; with these it is as we have said." (19a39)

That is, statements about the contingent future are true/false only in potency. And both, the affirmation and its negation are possibly true (even if most of the times, one is more true than the other). This is in accord with the thesis that they must have some truth-value.

But it is also the case that none of them *should be true* and the other false, for none of them is already true; that is, none of them is true in act. So *Tertio Excluso* does not apply to statements about future particulars for this would make their future necessary.

5. Other readings

I have pressed on the idea that Aristotle writes that statements about future contingents are *not yet* neither true nor false; and the idea that some divisions or composites are not *always* the case, to argue for the thesis that statements about future contingents are true/false in potency, but no in act. This also implies that they have truth-values in an important sense: they are potentially true.

My reading goes against the traditional/conventional reading, where Aristotle is thought to be defending that statements about future contingents are not true, nor are they false. They do lack truth-value. So *Tertio Excluso* does not apply universally. I agree that *Tertio Excluso* does not apply; but I disagree in the idea that they do not have truth/value: $18^{b}17$-18 seems to say that they must have.

There is, on the other side, Boethius's reading where he attributes to Aristotle the view that predictions are not "definitely true" or "definitely" false; and that this lack of definiteness is not a function of ignorance, but due to the nature of reality. I agree. But in Boethius's reading statements about future contingent must be either true or false (so *Tertio Excluso* applies). I disagree here. 19^{a}23-39 clearly says that future contingents are exceptions to it.

GOBERNANZA DE RIESGOS EN NANOMEDICINA: LA AMBIVALENCIA EUROPEA

HANNOT RODRÍGUEZ
Institute for Technology Assessment and Systems Analysis (Karlsruhe Institute of Technology), y Cátedra Sánchez-Mazas UPV/EHU
hannot.rodriguez@kit.edu

ANDONI IBARRA
Universidad del País Vasco (UPV/EHU)
andoni.ibarra@ehu.es

ABSTRACT: Nanomedicine is the application of nanotechnology to the study and improvement of human health, and a research area of high economic significance as well. But it also implies risks for human health and the environment, so the European Union claims to be committed to its "responsible development". However, there are different opinions in the EU about how nanomedicine—and nanotechnology in general—should be regulated. This paper will analyze critically two of the more relevant opinions.

KEYWORDS: nanomedicine, nanotechnology, risk, governance, Europe

Introducción

En el marco de las políticas reguladoras europeas sobre la nanotecnología, en general, y la nanomedicina, en particular, podemos identificar dos actitudes: una de tipo emprendedor, o pro-industria, representada por la Comisión Europea (CE); la otra cautelosa, o protectora de la salud y el medio ambiente, representada por el Parlamento Europeo (PE). Esta ambivalencia institucional pone de manifiesto la tensión derivada de promover de una manera segura esta tecnología mientras se trata de satisfacer el imperativo de la innovación, y ello en un contexto donde la ciencia ignora los mecanismos de toxicidad básicos de los nanomateriales. Esta contribución quiere dar cuenta de esta ambivalencia, y analizará: 1) las expectativas y preocupaciones generales asociadas al desarrollo de las nanotecnologías; 2) la importancia sanitario-industrial de las nanomedicinas; y 3) las características de riesgo de la nanomedicina, así como la respuesta reguladora de la UE a partir de las discrepancias entre la CE y el PE.

1. Nanotecnología: expectativas y riesgos

La nanotecnología es un nuevo *modo* de conocer y manipular la realidad con aplicaciones en prácticamente todos los sectores industriales de base tecnológica (e.g., medicina, energía, alimentación) (CE 2004, 4-6).

Este carácter capacitador, u "horizontal", de la nanotecnología, le confiere una gran importancia económica, y por ello la I+D en nanotecnologías ha despertado un gran interés tanto de las políticas científicas públicas como de la empresa privada (Berube 2006). Por ejemplo, ciertos análisis económicos, asumidos por la CE, predicen un mercado internacional valorado entre 750.000 millones y dos billones de euros, y la creación de 10 millones de empleos, ligados a la nanotecnología para el año 2015.[1]

Sin embargo, el desarrollo de las nanotecnologías también ha sido relacionado con la emergencia de nuevos riesgos sanitarios y medioambientales, ya que la manipulación de materiales a niveles molecular y atómico puede provocar riesgos relacionados con propiedades físico-estructurales que solo se producen a esos tamaños nanométricos (Maynard 2007). Por ello, la CE aboga por desarrollar la nanotecnología "de forma segura y responsable" (CE 2004, 3), lo que implica que el desarrollo de la nanotecnología debe descansar en una evaluación científica de sus riesgos que permita tomar medidas que garanticen la seguridad de las aplicaciones nanotecnológicas en la sociedad.

2. Nanomedicina: su relevancia sanitaria y socio-económica

La nanomedicina es la aplicación de la nanotecnología al estudio y la mejora de la salud humana, y promete mejoras en el diagnóstico médico y la terapia, así como en la regeneración de tejidos y órganos, por mor de las propiedades físico-químicas y biológicas de los nanomateriales. Por ejemplo, el uso de nanopartículas y puntos cuánticos permite el diagnóstico a niveles celular y sub-celular. En terapia, las nanopartículas posibilitan incidir directamente en la fuente de la enfermedad, lo que incrementa la eficiencia y minimiza los efectos colaterales asociados al uso de medicamentos no tan selectivos (Moghimi *et al* 2005). En lo tocante a la nanomedicina regenerativa, esta ha obtenido avances considerables en el desarrollo de piel, hueso y cartilago artificiales, y cuenta ya con una presencia parcial en el mercado.[2]

En este sentido, la nanomedicina es un elemento clave para la "Estrategia 2020" de la UE, que establece como uno de sus tres principales objetivos el desarrollo de un sistema de salud

[1] http://ec.europa.eu/nanotechnology/faq/faqs.cfm?lg=es&sub=details&idfaq=28821 (CE: "Preguntas más frecuentes sobre nanotecnología").
[2] http://www.etp-nanomedicine.eu/public/about/objectives-mission (ETP on NanoMedicine: "The Challenge").

asequible y de alta calidad para sus ciudadanos en una Europa cada vez más vieja. Así, la CE lleva invertidos ya, únicamente a través de su VIIPM (2007-2013), y durante los cuatro primeros años de este, más de 350 millones de euros en actividades de I+D relacionadas con la nanomedicina (Gabellieri y Frima 2011).

Además de su objetivo de mejora de la salud, la nanomedicina es considerada un elemento estratégico clave para la competitividad de la industria sanitaria en Europa (CE 2011). Así, la política científica europea promueve la integración de la industria en las actividades de I+D relacionadas con las diversas áreas de la nanotecnología: la integración de los actores industriales se considera un elemento sustancial en la evaluación de las propuestas de investigación que optan a financiación pública (e.g., CE 2007). Por otra parte, la propia industria, con el apoyo de la CE, ha ido adoptando iniciativas con el fin de optimizar su capacidad innovadora. Por ejemplo, en 2005 fue creada la Plataforma Tecnológica Europea sobre Nanomedicina, un consorcio liderado por la industria que engloba a organizaciones empresariales y académicas y a otros grupos de interés relacionados con el sector sanitario (ETPNM 2005). Esta plataforma estableció una Agenda de Investigación Estratégica con los objetivos de superar la fragmentación en la investigación nanomédica, obtener inversión pública y privada adicional, e identificar otras áreas prioritarias de investigación y desarrollo, entre otros (ETPNM 2005; 2006).

Aparte de la consecución de estos objetivos estratégicos, la PTE sobre Nanomedicina se hace eco del mandato de la CE de desarrollar esta tecnología “de forma responsable” (ETPNM 2005, 5). Esto se traduce en que el análisis de los riesgos sanitarios y medioambientales, el estudio de los aspectos éticos, sociales y legales (ELSA), así como la consideración de las opiniones de distintos grupos de interés y del público en general, son abordados como elementos constitutivos de los procesos de I+D en nanomedicina (ETPNM 2005; 2006).

La noción de “innovación responsable” se ha abierto paso en los ámbitos político y empresarial, motivado en parte por el interés en evitar resistencias sociales que, como la reciente experiencia del fiasco europeo de las biotecnologías agroalimentarias ha mostrado, pudieran cuestionar la explotación comercial de las nanotecnologías (Macnaghten 2008). La cuestión estriba ahora en observar el modo en que esta innovación responsable se articula con el principal objetivo de las políticas científicas y empresariales en el ámbito de la nanomedicina; a saber: la consecución de avances médicos con un alto valor de mercado (CE 2004; ETPNM 2005). Por ejemplo, únicamente un 2% (28 millones de euros) de la inversión total en nanotecnología (1400 millones) del VIPM de la CE fue dirigida a financiar proyectos de investigación cuyo fin era evaluar los riesgos sanitario-medioambientales asociados a las nanotecnologías (Aguar y Murcia Nicolás 2008,

7). A fin de cuentas, y aunque parezca obvio, el análisis de los riesgos es por definición una actividad secundaria que se aplica sobre otra actividad prioritaria: la innovación tecnológica de alta relevancia socio-económica.

3. Riesgos de la nanomedicina: entre la ambivalencia de la promoción y la regulación

La propia característica que hace de la nanomedicina tan interesante –i.e., el hecho de diagnosticar y curar a niveles de precisión desconocidos hasta ahora en virtud de su capacidad para conocer y manipular la realidad a niveles nanométricos– crea el potencial para nuevos riesgos sanitarios y medioambientales. El tamaño sí importa, y el comportamiento de los materiales a niveles nanométricos difiere del de sus equivalentes a escala macro, lo que implica que el comportamiento en términos de toxicidad también puede verse alterado en ciertos casos. Los nanomateriales, debido a su tamaño, traspasan las barreras biológicas más fácilmente, y además adquieren una mayor superficie en relación a su masa, lo que les confiere una mayor capacidad catalizadora pero a costa de un aumento en su toxicidad (Maynard 2007). Por ejemplo, algunos estudios han demostrado que la inhalación y aspiración de nanotubos de carbono que son utilizados como transportadores de fármacos provocan inflamación, formación de granuloma y fibrosis en el sistema pulmonar de ratas de laboratorio (Kagan *et al* 2005; Oberdörster 2010, 99-100). Otro ejemplo subrayaba que el cadmio y el plomo usados habitualmente como elementos de revestimiento de los puntos cuánticos, utilizados en procesos de bioimagen/ diagnóstico, tienden a degradarse en condiciones de luminosidad y oxidación, lo que podría provocar la translocación de estos metales hacia las células y otras partes del organismo, con efectos tóxicos (Hardman 2005).

De un modo genérico, se han identificado ya varias propiedades físico-químicas que determinarían la toxicidad de los nanomateriales:

Propiedades físico-químicas que determinan la toxicidad de las nanopartículas
Tamaño (considerando procesos de aglomeración/agregación en líquidos y aire)
Distribución del tamaño
Forma
Aglomeración/agregación

Propiedades de superficie
Superficie de área (incluyendo porosidad)
Carga
Reactividad
Composición química (incl. revestimientos, contaminantes)
Defectos
Solubilidad (lípida, acuosa, *in vivo*)
Cristalinidad

Fuente: Oberdörster 2010, 93.

A esta variedad de factores y a la complejidad derivada de sus múltiples combinaciones posibles en una amplia gama de nanomateriales, se le ha de sumar la complejidad añadida de que las propiedades pueden variar de acuerdo a los métodos de producción, los procesos de preparación y almacenamiento, así como su introducción en una heterogenidad de medios (ambiental, organismo, etc.) (Oberdörster 2010, 93).

Sin embargo, y a pesar de los avances en la comprensión y la identificación de estos factores de toxicidad, los mecanismos causales que gobiernan la emergencia de los efectos tóxicos permanecen en su mayoría inciertos (SCENIHR 2007). Es decir, los parámetros básicos en el marco de los cuales debería producirse nuevo conocimiento sobre la toxicología de los nanomateriales en su contacto con el organismo humano y el medio ambiente permanecen básicamente desconocidos. Existe una incertidumbre "radical" por ejemplo acerca del comportamiento de los nanomateriales en el medio, las rutas de exposición y la métrica por la cual debería medirse esta exposición, los mecanismos de translocación a otras partes del cuerpo, o los mecanismos exactos por los cuales se activan los procesos tóxicos (Kandlikar *et al* 2007). Esto implica que la propia posibilidad de desarrollar una evaluación científica cuantitiva fiable de los riesgos asociados a los nanomateriales se ve coartada, según reconocen los propios expertos que asesoran a la CE (SCENIHR 2009).

La manera de abordar este estado de ignorancia parcial acerca del verdadero alcance de los riesgos de los nanomateriales en el seno de la UE es diversa, y va desde una actitud de tipo "emprendedor" –que prioriza el desarrollo industrial de la nanotecnología sobre la seguridad– hasta una de tipo "cauteloso" –que confiere un mayor peso a las medidas de seguridad frente a la promoción de las nanotecnologías.

En el lado emprendedor, podemos enumerar una variedad de iniciativas protagonizadas por la CE. Por ejemplo, esta concluyó en un informe de 2008 que el marco regulador europeo pre-nanotecnológico –basado por tanto en criterios de seguridad anteriores a los desafíos planteados por

los nanomateriales y sus propiedades novedosas– era apropiado para regular los riesgos de las nanotecnologías: *"Current legislation covers in principle the potential health, safety and environmental risks in relation to nanomaterials"* (CE 2008a, 11). Es posible que esta conclusión se apoyara, entre otras, en la opinión previa del Comité para los Productos Medicinales de Uso Humano de la Agencia Europea sobre Medicinas, que en 2006 concluyó la validez del marco regulador sobre productos y dispositivos medicinales "convencionales" para tratar con los nuevos riesgos de las nanomedicinas: *"The evaluation and prevention of potential hazards related to the use of any given "nanomedicinal" products is already foreseen under the existing EU pharmaceutical legislation"* (citado en: Jahnel *et al* 2012). Fiel a este desinterés por el desarrollo de normas vinculantes que pudieran obstaculizar el desarrollo de las nanotecnologías, la CE publicaba en 2008 un código de buena conducta voluntario, dirigido a los sectores relacionados de un modo u otro con la nanotecnología (empresarios, científicos, políticos, ONGs, etc.) (CE 2008b).

En el lado de la cautela se ubica el PE, la única institución europea elegida directa y democráticamente por los ciudadanos y, en consecuencia, la más proclive a defender los intereses de los consumidores y el medio ambiente por encima de los intereses empresariales. En este sentido, el PE se ha mostrado crítico con la CE y ha demandado una nueva investigación al ejecutivo acerca de la validez del marco normativo europeo para regular las nanotecnologías, además de hacer valer su capacidad legisladora para implicarse en el desarrollo y aprobación de regulaciones que incluyen normas diseñadas específicamente para la regulación de los nanoproductos (Rodríguez, en prensa). De entre estas regulaciones la más relevante –en términos de la especialización y variedad de sus normas– es la regulación 1223/2009 sobre productos cosméticos, que incluye una amplia gama de normas que demandan, por ejemplo, una evaluación científica especial de los nanomateriales, o el etiquetado de todo producto cosmético que contenga nanomateriales (The European Parliament and the Council of the EU 2009). De este modo, la UE se ha convertido en el primer gobierno (a nivel tanto nacional como transnacional) que desarrolla regulaciones específicamente dirigidas a las nanotecnologías. Sin embargo, otra serie de propuestas más "radicales" del PE han sido rechazadas de plano por la CE (y por la otra rama legisladora de la UE junto al PE, el Consejo Europeo), como por ejemplo su demanda, basada en una interpretación fuerte del principio de precaución, de aplicar una moratoria de facto a los nanoalimentos en la propuesta de normativa europea sobre nuevos alimentos hasta que se desarrollen métodos de evaluación científica adecuados de los nanoriesgos y se reduzcan las incertidumbres científicas en torno a ellos (Rodríguez, en prensa).

Por supuesto, tanto la CE como el PE tienen interés en promover las nanotecnologías, y de un modo responsable, o seguro, por lo que sería demasiado simplista identificar a cada una de estas instituciones con una actitud unidimensional respecto al riesgo (i.e., CE: emprendedora; PE: cautelosa). Sin embargo, este ejemplo pone de manifiesto esta tensión esencial entre promoción y control que caracteriza a las políticas de ciencia comprometidas con el desarrollo seguro de las tecnologías. En un contexto internacional donde los países industrializados invierten cada vez más dinero y esfuerzo en la innovación (Marklund *et al* 2009), una sobrerregulación unilateral de las nanotecnologías supondría un contratiempo sensible para los intereses económicos de la industria europea –si bien una infrarregulación (en tanto así percibida por la sociedad) podría provocar un efecto similar (e.g., el caso de la biotecnología agroalimentaria). Por ejemplo, queda por ver el alcance de la tolerancia política para con las incertidumbres científicas que rodean a los riesgos de las nanotecnologías, y la nanomedicina en particular, y si la aplicación del principio de precaución pudiera conducir a una moratoria sobre nanomateriales, o a una actitud institucional menos optimista con respecto a las capacidades cientificas de conocimiento y reducción de complejidad, implícitas en la idea de "desarrollo responsable".

Agradecimientos

La contribución de Hannot Rodríguez a este trabajo ha sido posible gracias a una beca del *Karlsruhe Institute of Technology* (KIT) de Alemania (1/5/2012-30/4/2013), y además se ha realizado en el marco del proyecto de investigación KONTUZ! del Ministerio de Economía y Competitividad: "Los límites del principio de precaución en la praxis ético-jurídica contemporánea" (ref. FFI2011-24414).

REFERENCIAS

Aguar, P., Murcia Nicolás, J.J. 2008, *EU nanotechnology R&D in the field of health and environmental impact of nanoparticles*, Bruselas, EC, Research DG.

Berube, D.M. 2006, *Nano-Hype: The Truth Behind the Nanotechnology Buzz*. Amherst, NY, Prometheus Books.

CE 2004, *Hacia una estrategia europea para las nanotecnologías*, Bruselas, COM(2004) 338.

CE 2007, *Work Programme 2007, Cooperation, Theme 4, Nanosciences, Nanotechnologies, Materials and New Production Technologies,* C(2007)560.

CE 2008a, *Regulatory Aspects of Nanomaterials*, Bruselas, COM(2008) 366.

CE 2008b, *Code of conduct for responsible nanosciences and nanotechnologies research*, Bruselas, C(2008) 424.

CE 2011, *Targeted Nano-Pharmaceuticals and Early Dianostics Cluster*, R&I / IT.

ETPNM, ETP on NanoMedicine 2005, *Vision Paper and Basis for a Strategic Research Agenda for NanoMedicine*. Luxemburgo, OOPEC.

ETPNM 2006, *Nanotechnology for Health – Strategic Research Agenda*. Luxemburgo, OOPEC.

The European Parliament and the Council of the EU 2009, "Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products', *Official Journal of the European Union*, L 342/59, 22.12.2009.

Gabellieri, C.; Frima, H. 2011, 'Nanomedicine in the European Commission policy for Nanotechnology', *Nanomedicine: Nanotechnology, Biology, and Medicine* 7(5), 519-20.

Hardman, R. 2005, 'A toxicological review of quantum dots: Toxicity depends on physico-chemical and environmental factors', *Environmental Health Perspectives* 114(2), 165-72.

Jahnel, J.; Fleischer, T.; Seitz, S. 2012, 'Challenges and Questions in Nanomedicine Legislation', Spring Conference "Nanomedicine: Visions, risks, potential". Berlin-Brandenburg, Academy of Sciences and Humanities, 19-20/4/2012.

Kagan, V.E.; Bayir, H.; Shvedova, A.A. 2005, 'Nanomedicine and nanotoxicology: two sides of the same coin', *Nanotechnology: Nanomedicine, Biology, and Medicine* 1(4), 313-16.

Kandlikar, M.; Ramachandran, G.; Maynard, A.; Murdock, B.; Toscano, W.A. 2007, 'Health risk assessment for nanoparticles: A case for using expert judgment', *Journal of Nanoparticle Research* 9(1), 137-56.

Macnaghten, P. 2008, 'From Bio to Nano: Learning the Lessons, Interrogating the Comparisons', in David, K. y Thompson, P.B. (eds.), *What Can Nanotechnology Learn From Biotechnology?*, Amsterdam, Academic Press, 107-23.

Marklund, G.; Vonortas, N.S.; Wessner, C.W. (eds.) 2009, *The Innovation Imperative: National Innovation Strategies in the Global Economy*, Cheltenham, UK, Edward Elgar.

Maynard, A.D. 2007, 'Nanotechnology: The Next Big Thing, or Much Ado about Nothing?', *Annals of Occupational Hygiene* 51(1), 1-12.

Moghimi, S.M.; Hunter, A.C.; Murray, J.C. 2005, 'Nanomedicine: current status and future prospects', *The FASEB Journal* 19(3), 311-30.

Oberdörster, G. 2010, 'Safety assessment for nanotechnology and nanomedicine: concepts of nanotoxicology', *Journal of Internal Medicine* 267(1), 89-105.

Rodríguez, H. (En prensa), 'Nanotechnology and Risk Governance in the European Union: The Constitution of Safety in Highly Promoted and Contested Innovation Areas', in Miller, C., Barben, D. y Molfino, E. (eds.), *The Yearbook of Nanotechnology in Society. Volume 4: Nanotechnology and Democracy*, Nueva York, Springer.

Scenihr, Scientific Committee on Emerging and Newly Identified Health Risks 2007, *Opinion on the Appropriateness of the Risk Assessment Methodology in Accordance with the Technical Guidance Documents for New and Existing Substances for Assessing the Risks of Nanomaterials*, Bruselas, EC, Health & Consumer Protection, DG.

Scenihr 2009, *Risk Assessment of Products of Nanotechnologies*, Bruselas, EC, DG for Health & Consumers.

EXPERIMENTS IN THE SOCIAL SCIENCES: RETHINKING THE HAWTHORNE EFFECT

MARIA JIMENEZ-BUEDO
UNED

ABSTRACT: The paper focuses on a particular undertheorized phenomenon in the experimentation of the social sciences: the Hawthorne effect, generally defined as the problem in experiments whereby subjects' knowledge that they are in an experiment modifies their behaviour from what it would have been without the awareness of being an experimental subject. Although the Hawthorne effect is often mentioned in the methodological discussions of social science experimenters and in virtually all research design textbooks, the notion is often unsatisfactorily depicted, where contradictory definitions can be found in the literature. The paper reviews the standard definitions and common methodological strategies of experimenters for dealing with this type of 'reactivity'. We focus, in particular, on a recent attempt by Zizzo to define and deal with this phenomenon. The paper shows the limitations of his conceptual strategy and reflects on the characteristics that an alternative definition should have in order to be useful for social science experimenters.
KEYWORDS*:* Hawthorn effect; reactivity; experiments; experimental economics.

In the last years, and particularly since the late 1990s, the number of social science articles vindicating the experimental method as the ideal means both for the generation of primary data and of empirical evidence in support for a given theoretical stance has increased notably, even in some social sciences in which its use was traditionally thought of as difficult or even impracticable, such as sociological theory, political science, and above all, economics (Levitt and List, 2007; Morton and Williams, 2009). The revival of experimentation in those social scientific disciplines that lack a long standing experimental tradition (in particular, economics) is yet to be met by a subsequent effort on the part of methodologists and philosophers of science in order to adapt existing theoretical accounts of experimentation to the terrain of the social sciences. Although exceptions do exist (see, for example, Guala 2005), many of the methodological discussions around experiments in economics are framed with the terms and accounts that were developed attending to the characteristics of the natural sciences and their patterns of experimentation.

In particular, the commonly held picture is that experiments in social science are best understood by reference to the well-known structure of controlled variation of the putative cause; and that the best way to capture the effect of the putative causes is to calculate the difference in means in the variable of interest of the treatment and control groups. However, and as it will be argued in this paper, many experiments in behavioural economics, partially following on a long-

standing tradition in social psychology, do not follow that structure, but instead, are best defined by the experimenter's recreation of a complex scenery (i.e., an artificial environment) that is meant to let the inner capacities and dispositions of the experimental subject emerge, in a setting that is nothing alike the idea of controlled variation of potential causal elements one by one. Although the majority of experimenters in the area of behavioural economics do follow the "controlled variation" scheme in the presentation of their experimental results, this logic does not always properly account for the design of the settings with which they present their subjects. This mismatch between the notion of an ideal experimental design and the actual structure of many experiments in the social sciences raises questions, in particular, regarding the reactivity of experimental social science subjects.

The paper focuses on a particular undertheorized phenomenon in the experimentation of the social sciences: the Hawthorne effect, generally defined as the problem in experiments whereby subjects' knowledge that they are in an experiment modifies their behaviour from what it would have been without the awareness of being an experimental subject. Although the Hawthorne effect is often mentioned in the methodological discussions of social science experimenters and in virtually all research design textbooks, the notion is often unsatisfactorily depicted, where contradictory definitions can be found in the literature. We review the common standard methodological strategies of experimenters for dealing with 'reactivity', as the Hawthorne effect is often referred to (Adair, 1984; Chiesa and Hobbs, 2008). The paper's position is thus that the Hawthorne effect is something more than a methodological annoyance, and that it needs to be addressed by an epistemology of social science experimentation that is now on the making.

As a case study the paper examines the controversy that initiated mainly around Bardsley's results (Bardsley 2005, 2008), whereby, in an innovative variant of the Dictator Game, subjects were found to take money from recipients when given the choice, thus putting into question the idea that standard results in DGs constituted evidence of altruistic preferences. Equally, Dana et al (2007) had shown that subjects in the dictator condition preferred to remain ignorant regarding the relationship between their actions and resulting outcomes so that they can be able to behave self-interestedly without appearing "unfair", either to themselves or others. In view of this, Daniel Zizzo (2010) has argued that the results of Dictator Games, based on highly artificial settings, are particularly prone to what he has coined as "experimenter demand effects", defined as changes in behaviour by experimental subjects due to unintended cues provided by the experimenter (or the

experimental setting) about what constitutes appropriate behaviour. Thus, accordingly, suggests that DG designs have been abused in experimental economics and are best abandoned.

Zizzo's position, while widely cited and influencing, has not been the only reaction possible to the results of Bardsley's and others. Guala and Mittone (2010), in a piece with the same publication date but written previously to that of Zizzo's also take issue with the controversy regarding the standard interpretation of DG results. They, however, defend the idea that the DG is potentially a very useful tool for experimental game theory in the study of social norms, and argue, instead, that what is lacking among experimental economists is an adequate theory of norms, which they think should be inspired by recent work on the matter by Bicchieri (2006) among others. Guala and Mittone argue that the DG is useful, precisely because it being an unusual and abstract design it is particularly suitable for the introduction of contextual cues that determine the framing and triggering of social norms, which is, in their view, the way to go about testing theories concerning norms. This line of defence regarding the DG has been subsequently addressed and opposed by Zizzo (2011), insisting in that this game is particularly badly suited to the study of social norms because it is especially prone to being subject to experimenter demand effects.

The paper argues that while Zizzo's indictment of the DG is not unfounded, it ultimately rests on a conceptualization of "experimenter demand effects" that is problematic, in that his position, if unqualified, seems to suggest that behavioural and experimental economics are not well suited to the study of social norms. Against this idea, this paper analyzes critically Zizzo's conceptualization of the problem of experimenter demand effects.

First, we trace back the origins of the term to the social psychology literature where the phenomenon of reactivity of experimental subjects was first analyzed. The current interpretations and conceptualizations of experimenter effects and related phenomena were mainly coined by social psychologists (Orne, Rosenthal, Kruglanski) or by researchers refining the concept of Hawthorne effects, stemming from field tryouts in industrial and organizational psychology (see Adair 1984, Chiesa and Hobbs 2008). Though the terms *demand effect of experimentation*, *experimenter effects*, and the more widespread *Hawthorne effects* refer to related phenomena, there are distinct definitions to each of them, however, they are often (and confusingly) used interchangeably.

Second, we take issue with Zizzo's definition of the term experimenter demand effects. Our criticism of his conceptualization is based on the idea that Zizzo's concepts conflate two distinct

phenomena: namely, experimenter effects and demand effects of experimentation. Regarding the latter Orne (1962, 1967) argued that the experimental setting is a socially ambiguous scenario where a "demand effect of experiments" is likely to take place despite the experimenter's efforts to control for his or her influence. By this, Orne identified the phenomenon through which the experimental subject would have a motivation in guessing what the experiment 'was about', or to come up with a candidate for the 'experimental hypothesis', often to comply with it and 'be a good experimental subject'. Thus for Orne, the inherent ambiguity in the experimental setting, in terms of the type of social setting that it represents, creates an expectation in the experimental subject upon which he or she reacts. In contrast, Rosenthal (1967) coined the term of "experimenter effects" and empirically showed how experimenters recurrently communicated (both covertly and unconsciously) their expectations about the experiments' results and how this, in turn, impacted the experimental subjects' behaviour thus biasing results in favour of the experimental hypothesis. We argue that there is no theoretical advantage to the conceptual synthesis of these two phenomena offered by Zizzo: conflating these two phenomena simply obscures a useful distinction and ultimately masks a tension in the kind of methodological advice that is to be followed if either experimenter of demand effects of experimentation want to be controlled for, so for example: detailed instructions seem to have more potential for creating problematic experimenter effects, while ambiguous or incomplete instructions open the room for demand effects of experimentation.

In light of a thorough analysis of Zizzo's and alternative conceptualizations of the phenomenon of the reactivity of experimental subjects in behavioural and experimental economics, the paper concludes by calling attention to the centrality of the problem of reactivity of experimental subjects, an issue that is central to the notions of validity most commonly referred to in the discipline (be these internal, external or construct validity), and that yet lacks a conceptualization over which experimenters can agree.

REFERENCES

Adair, J. G. 1984, 'The Hawthorne Effect: A Reconsideration of the Methodological Artifact', in *Journal of Applied Psychology*. Vol 69, 2.

Bicchieri 2006, *The Grammar of Society: The Nature and Dynamics of Social Norms,* Cambridge University Press.

Cartwright, N. 2006, *Well-Ordered Science: Evidence for Use,* Philosophy of Science.

Chiesa, M. And H. Hobbs.(2008) Making Sense of Social Research: How useful is the Hawthorne effect? *European Journal of Social Psychology*, 38. 67-74.

Levitt, J. and J.A. List. 2009, 'Was there Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments' in *NBER Working Paper* No. 15016

Bardsley, N. 2005, 'Experimental Economics and the Artificiality of Alteration.' in *Journal of Economic Methodology* 12 (2005), 239-251.

_____2008, "Dictator Game Giving: Altruism or Artefact?" in *Experimental Economics* 11, no. 2, 122-133.

Dana, J., R. Weber, and J. K. Kuang 2007, "Exploiting Moral Wriggle Room: Experiments Demonstrating an Illusory Preference for Fairness." in *Economic Theory* 33, 67-80.

Guala, F. 2005, *The methodology of experimental economics*, Cambridge, Cambridge University Press.

Kruglanski, A. 1975, "The Human Subject in the Psychology Experiment: Fact and Artifact." in *Advances in Experimental Social Psychology*, ed. L. Berkowitz, 8, 101-147. N. York: Academic Press

Levitt, S.D. and J.A. List 2007 'What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?' in *Journal of Economic Perspectives*. Vol 21, 2, 153-174.

Morton, R.B. y K. C. Williams 2008, 'Experimentation in Political Science' in Janet Box-Steffensmeier, David Collier y Henry Brady. *The Oxford Handbook of Political Methodology,* Oxford, OUP

Orne, M.T. 1962, 'On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications.' in *American Psychologist* 17, 776-783.

Rosenthal, R., 1967, 'Covert Communication in the Psychological Experiment.' in *Psychological Bulletin* 67, 5, 356-367.

Trochim,W.M.K., *Advances in Quasi-Experimental Design and Analysis*, San Francisco, Jossey-Bass.

Zizzo, D. 2010, 'Experimenter Demand Effects in Economic Experiments.', in *Experimental Economics* 13, 1, 75-98.

Zizzo, D. 2011, 'Do Dictator Games Measure Altruism?' in *Handbook on the Economics of Philantrhopy, Reciprocity and Social Entreprise*. Bruni, L. and Zamagni, S. (eds.)

LA EXPLICACIÓN EN BIOLOGÍA: ENTRE LEYES Y MODELOS EXPLICATIVOS

VÍCTOR J. LUQUE MARTÍN
Universitat de València

ABSTRACT: The lack of consensus on the existence of empirical laws in biology, shows that biologist have been dedicating their time to make (essentially mathematical) models to express biological facts. The challenge is to find out how models explain. Thus, this paper analyzes Alisa Bokulich's explanatory model known as model explanation, and shows a specific example in biology with the Hardy-Weinberg equilibrium model.
KEYWORDS: biology, laws, explanatory models, model explanation, Hardy-Weinberg equilibrium

1. El problema de las leyes

Durante años, la polémica sobre la existencia de leyes en biología ha preocupado y enfrentado a los filósofos de la ciencia. Su existencia o no era de suma importancia puesto que la noción de explicación científica de origen hempeliano, exigía la presencia de–al menos– una ley en el *explanans* para considerarla como tal[1].

Aunque la polémica continua hoy en día, con posiciones tanto a favor de la existencia de leyes (cf. Elgin 2006) como en contra (cf. Beatty 1995) ello no es óbice para constatar que, si finalmente encontramos leyes empíricas en la biología, los biólogos han estado trabajando sin ellas durante los últimos ciento cincuenta años. La posición de Elliott Sober apunta en este sentido. Sober defiende que hay una fuerte similitud entre la función de las leyes en física y los modelos matemáticos en biología (Sober 2000). Ambos tienen la forma característica *si/entonces*, donde si las condiciones preliminares se dan en un espacio y tiempo concretos, entonces se cumplirán las consecuencias postuladas por la ley o el modelo. La ley de la gravedad de Newton postula que dos cuerpos se atraen con una fuerza directamente proporcional al producto de sus masas, e inversamente proporcional al cuadrado de la distancia que los separa. De la misma manera, el modelo de proporción sexual de Fisher postula que si en una población hay apareamiento al azar y

[1] Para una crítica clásica a la falta de leyes en la biología véase Smart (1963). Para una pronta defensa de los ataques de Smart, véase Ruse (1979). J.J.C. Smart defendía que ninguna de la leyes de la biología cumplían los dos requisitos (tradicionales) que debe tener una ley: 1) la «universalidad no restringida» (la ley debe aplicarse verdaderamente en todo el universo y en todo tiempo); y 2) la ley no debe especificar ningún lugar u objetos concretos. Por su parte, Michael Ruse constata que los físicos continúan hablando de leyes a pesar de no cumplirse el primer requisito en algunos casos (por ejemplo, la ley de Boyle es falsa a altas temperaturas y presiones) ni el segundo en otros (como las leyes de Kepler). Entonces, las llamadas leyes en biología se pueden considerar como tales sin perjuicio alguno.

las parejas parentales difieren en la combinación de hijos machos y hembras que producen (siendo esta diferencia heredable), entonces la proporción de sexos evolucionará hacia una proporción de 1:1 y se estabilizará. Ni el primero nos dice cuántos cuerpos hay realmente en el Universo, ni el segundo cuántos seres vivos de reproducción sexual existen. Simplemente, si se dan las condiciones demandas, tanto la ley de Newton como el modelo de Fisher se cumplirán.

La intuición de Sober es la que nos parece que hay que seguir. Efectivamente, los biólogos se han dedicado, no a buscar leyes sino a formular modelos (esencialmente matemáticos) que les permitieran operar con la cantidad ingente de datos en los que están insertados los seres vivos. Si son los modelos el campo de batalla de los biólogos entonces, ¿explican la realidad o no? ¿Cómo puede explicar un *modelo*, cuyo principal rasgo es la simplificación, la idealización e, incluso, la ficcionalización? Nuestra respuesta es que ciertos modelos tienen poder explicativo por sí mismos. El reto es desentrañar cómo éstos, de hecho, pueden explicar.

2. Modelos explicativos: la *explicación modelo* de Bokulich

Alisa Bokulich, en un reciente artículo titulado 'How Scientific Models Can Explain' (2011) analiza tres diferentes aproximaciones que se ha hecho en la literatura filosófica con respecto a la capacidad explicativa de los modelos científicos: el *mechanistic model explanation* de Carl Craver, el *covering-law model explanation* de Mehmet Elgin y Elliott Sober[2], y el *causal model explanation* de Ernan McMullin[3].

Para Craver (2006), el poder explicativo de un modelo refiere a cuando describe el comportamiento de todo el sistema en términos mecánicos (de su operación e interacción). Para ello, se necesitaría una caracterización completa del fenómeno que se pretende explicar, pero Bokulich rechaza dicha posibilidad porque tendríamos una descripción teorética del sistema y no un modelo, puesto que se habrían perdido las características principales de éste: la no literalidad de aquello que se pretende explicar mediante simplificaciones, idealizaciones o ficcionalizaciones.

Elgin y Sober (2002), por su parte, defienden que un modelo es explicativo (a pesar de sus idealizaciones) cuando el *explanans* describe la(s) causa(s) del *explanandum*, además de citar una ley de la naturaleza, y, por último, todas las proposiciones del *explanans* sean ciertas, en tanto que las idealizaciones del modelo sean "inofensivas" (*harmless*). El modelo, a pesar de permitir idealizaciones, tiene el lastre de la necesidad de que los enunciados que constituyen el *explanans*

[2] No confundir con el *covering law model of explanation* hempeliano.
[3] Los nombres de los modelos son denominaciones de la misma Bokulich (McMullin, como veremos, denomina al suyo *hypothetico-structural*).

sean verdaderos, lo que provoca (según Bokulich) que elimine precisamente la fuerza explicativa del modelo al contener idealizaciones o ficcionalizaciones.

Por último, McMullin (1978 y 1985) defiende un modelo explicativo que denomina *hypothetico-structural*, donde se alude a la estructura de la que está compuesta una entidad compleja para explicar sus propiedades o su comportamiento. Dicha descripción es, nos dice McMullin, "causal, since the structure invoked to explain can also be called the *cause* of the feature being explained" (McMullin 1978, 139)[4]. Pero además, al ser ésta originariamente de carácter hipotético o metafórico, debe permitirnos un proceso de desidealización (*de-idealization*) no como mero ajuste *ad hoc* del modelo a los datos, sino cuando haya una justificación teórica que lo permita. Este modelo, que es por el que siente más simpatía Bokulich, arrastra el problema de no admitir en su seno modelos que no se basan en idealizaciones sino en ficcionalizaciones (como es, para Bokulich, el modelo atómico de Bohr) [5], ni ámbitos científicos donde no existen las explicaciones causales (como en la mecánica cuántica).

Bokulich, por su parte, va a defender un modelo propio, con el nombre de *model explanation*, donde los modelos explicativos anteriores son en realidad casos especiales. En sus propias palabras, un modelo explica cuando:

> "First, the explanans makes reference to an idealized or fictional model; Second, that model explains the explanandum by showing that the counterfactual structure of the model is isomorphic (in the relevant respects) to the counterfactual structure of the phenomenon[6]. This means that the model is able to answer a wide range of "what-if-things-had-been-different" questions. And third, there is a justificatory step specifying what the domain of applicability of the model is and that the model is an adequate guide to that domain of phenomena"[7] (Bokulich 2011, 43).

Dicho modelo abarca los anteriores en tanto que permite identificar qué tipo de modelo son mostrando su origen de dependencia: si producen causalmente el *explanandum* (es decir, el hecho)

[4] Por ejemplo, la causa de la emisión de la energía radiante del Sol es la fusión del hidrógeno, transformándose en helio, que se produce en el interior de éste.

[5] En los últimos años se ha producido un auge en la literatura filosófica que considera a los modelos científicos como *ficciones*. El término es problemático en cuanto parece equiparar, pudiendo llevar a equívocos, las ficciones literarias a las científicas. La diferencia, brevemente, entre un modelo idealizado y uno ficcional es que el primero permite un proceso de desidealización, de aproximaciones a la realidad, mientras que el otro no. De este modo, el modelo atómico de Bohr sería ficcional porque no es posible ningún tipo de desidealización o corrección, en tanto que hoy sabemos que el movimiento de los electrones se produce en una nube probabilística de densidad y no en trayectorias discretas circulares. Para una aproximación reciente al problema de la noción de *ficción* en ciencia, véase Iranzo (2011).

[6] Debe ser isomórfica en el sentido relevante para la estructura contrafáctica del fenómeno, reproduciendo el modelo, en un sentido relajado (*loose sense*), las características relevantes del *explanandum*.

[7] Dicho paso justificativo puede producirse de dos maneras: de forma descendente (*top down*) desde la teoría, donde nos especifica los dominios fiables del modelo; o de forma ascendente (*from the ground up*) a través de la investigación empírica.

se tratará del modelo causal de McMullin, si lo hacen siendo las partes mecánicas se tratará del modelo mecanístico de Craver, y si se debe a las leyes citadas en el modelo se tratará del propuesto por Elgin y Sober. Pero además permite un subtipo más, que Bokulich denomina *structural model explanation*, el cual podemos encontrar en la física (en campos como la relatividad o la mecánica cuántica) y que consiste, no en explicaciones causales sino en la dependencia estructural que se produce entre la teoría (o teorías) dentro del modelo y el *explanandum*, limitando los objetos, propiedades, estados, etc., que encajarían en ella(s).

Aunque Bokulich se centra (al ser su campo de especialidad) en los modelos de la mecánica cuántica, nosotros queremos mostrar que dicho modelo es aplicable a, al menos, algunos modelos de la biología (esencialmente matemáticos). Nos centraremos en el modelo de equilibrio Hardy-Weinberg (de forma abreviada equilibrio H-W), aunque, si estamos en lo correcto, la *explicación modelo* de Bokulich puede aplicarse a un gran número de modelos matemáticos en biología, no sólo a la H-W sino a otros como el modelo de proporción sexual de Fisher, el de depredador/presa de Lotka-Volterra, el modelo de teoría evolutiva de juegos de Maynard-Smith, etc.

3. Equilibrio Hardy-Weinberg

En 1908, y de forma independiente, el matemático inglés G. H. Hardy y el fisiólogo alemán Wilhem Weinberg, desarrollaron un modelo matemático que permitía introducir las matemáticas en el mundo de la biología. El modelo de equilibrio H-W postula que en una población idealmente infinita, donde hay apareamiento al azar (población panmíctica) y cuyos individuos son tanto viables como fecundos, la población se mantendrá o volverá al equilibrio si no actúa ninguna fuerza sobre ella, puesto que la herencia mendeliana por sí sola no puede cambiar las frecuencias alélicas de aquélla. La demostración en su forma más simple se realiza así: con dos alelos en un mismo locus, A y a, y las frecuencias de los cuales son p y q respectivamente, entonces las frecuencias de los tres genotipos AA, Aa y aa serán p^2, $2pq$ y q^2 respectivamente.

Pero dicho modelo permite formulaciones más complicadas. Ya sea con alelos múltiples, para un gen ligado al sexo (puesto que la formulación original solo era para genes autosómicos), poliploidía (más de dos juegos de cromosomas) o equilibrio para dos *locis*. Por mor de la brevedad, sólo demostraremos uno de ellos: el gen ligado al sexo. De la formulación cromosómica, XX/XY, tenemos cinco genotipos posibles: AA, Aa y aa en hembras, y A y a en machos. Siendo p la frecuencia de A y q la de a, las frecuencias genotípicas en equilibrio son $p^2(AA)$, $2pq(Aa)$, y $q^2(aa)$ para las hembras, y $p(A)$ y $q(a)$ para los machos, coincidiendo las frecuencias alélicas con las genotípicas (Cabrero & Camacho 2003, 85).

Vemos que este modelo clásico de equilibrio H-W cumple los requisitos del modelo explicativo propuesto por Bokulich. El primer punto es satisfecho al ser un modelo idealizado puesto que postula una población infinita, donde hay apareamiento al azar y los individuos son tantos viables como fecundos. Los biólogos saben perfectamente que no hay poblaciones infinitas y que, muy raramente, todos los individuos son viables y fecundos. Pero dichas idealizaciones les permiten manejar el inmenso conjunto de datos y variables en los que están insertados los seres vivos.

El segundo punto es satisfecho en tanto que el modelo de equilibrio H-W responde a preguntas del tipo *what-if-things-had-been-different*. La primera formulación se basó en genes autosómicos, pero hemos mostrado (y demostrado uno de ellos) que el modelo es capaz de responder a situaciones modificadas: genes ligados al sexo, alelos múltiples, etc. Por supuesto, el modelo no permite infinitas correcciones ni desviaciones de sus presupuestos, pero ese no es su cometido ya que, si así fuera, no sería más que un modelo fenomenológico aplicable a cualquier situación y cambio. Ello lo haría útil para el científico, pero no le daría una explicación y entendimiento mejor del mundo. Lo que el modelo H-W nos da, a través de su resultado de estabilidad de las poblaciones sexuales, es una información crucial de las poblaciones mendelianas, de cómo se produce la herencia y la variación.

En cuanto al último requisito, el dominio de aplicación del modelo H-W queda claro gracias a la Genética de Poblaciones[8], mostrándose adecuado para poblaciones de gran tamaño pero no para las de pequeño tamaño, donde la deriva genética produce desvíos al azar de las frecuencias génicas respecto de las frecuencias genotípicas esperadas[9].

4. Conclusiones: problemas pendientes

A pesar del atractivo del modelo explicativo presentado, Alisa Bokulich no deja del todo claros diversos puntos de su propuesta. El requisito de isomorfía entre las estructuras contrafácticas

[8] McMullin subraya que una de las características fundamentales de los modelos idealizados es que son (o suelen ser) la base para el surgimiento de nuevos programas de investigación (entendidos de forma mucho más modesta y restrictiva que por Lakatos) en cuanto propician nuevas investigaciones que aumentan el número de explicaciones del fenómeno explicado en un primer momento, y un proceso sucesivo de desidealización del primer modelo. Así, el modelo de equilibrio Hardy-Weinberg fue la base del surgimiento de una de las áreas más potentes y matematizadas de la biología: la Genética de Poblaciones. De hecho, gracias a ella se produjo la famosa Síntesis Moderna, donde la genética de raíz mendeliana y la selección natural darwiniana se fundían en un nuevo marco. Esto confirmaría una de las intuiciones de Bokulich (en tanto que piensa que el *paso justificativo* la mayoría de las veces es una combinación de las dos formas) ya que se dio de las dos formas posibles conjuntas, solo que en diferentes periodos de tiempo (primero de forma ascendente y, hoy en día, de forma descendente).

[9] De forma más técnica, la deriva genética es un: "Proceso que, debido a la finitud de tamaño de las poblaciones, genera desviaciones estadísticas de las frecuencias génicas en la muestra de gametos de la descendencia, respecto a las esperadas por la segregación mendeliana" (Fontdevila & Moya 2003, 21).

del modelo y el fenómeno, parece dejar en el aire el mismo concepto de isomorfía. A pesar de ésta exigir que, dados dos sistemas, sus estructuras sean idénticas, Bokulich precisa que el modelo debe reproducir el *explanandum* de forma aproximada, en un sentido laxo (*loose sense*). La isomorfía es una noción estructural fuerte a la que Bokulich no parece estar dispuesta a renunciar, ya que podría haber hablado de una simple relación de semejanza entre la estructura del modelo y el fenómeno.

A su vez, su *paso justificativo* podría, en vez de aclararlo, emborronar el lugar donde descansa el poder explicativo de los modelos. Bokulich plantea la participación del entramado teórico para asegurar que los modelos no sean meros modelos fenomenológicos que salven los fenómenos. Si el modelo no fuera, finalmente, más que una "aproximación práctica" de la teoría (o conjunto de teorías), entonces no habría necesidad de desentrañar ningún rol explicativo por parte del modelo, puesto que sería la teoría (con sus leyes o regularidades más o menos abarcantes) la que jugaría tal rol.

No hay duda de que los modelos son una parte central del quehacer de los científicos (y en el caso de la biología aún más si cabe), y puede que fuera mejor, antes de buscar un modelo explicativo omniabarcante, realizar una taxonomía de aquellos antes de desentrañar su poder explicativo[10].

Los problemas en las caracterizaciones anteriores, tanto los de los autores tratados por Bokulich como el suyo propio, nos muestran la dificultad y el camino que aún nos queda por recorrer en la aprehensión explicativa de los modelos. No obstante, el modelo explicativo propuesto por Alisa Bokulich nos parece un intento interesante que puede conducirnos a una conceptualización (o conceptualizaciones) más precisa de cómo los modelos explican, con su importancia en la explicación biológica.

REFERENCIAS

Beatty, J. 1995, 'The Evolutionary Contingency Thesis', in Wolters, G. and Lennox, J. G. (eds.), *Concepts, Theories, and Rationality in the Biological Sciences*, The Second Pittsburgh-Konstanz Colloquium in the Philosophy of Science, Pittsburgh, University of Pittsburgh Press, 45-81.

Bokulich, A. 2011, 'How Scientific Models Can Explain', *Synthese*, 180 (1), 33-45.

Cabrero, J. & Camacho, J. P. 2003, 'Fundamentos de Genética de Poblaciones', en Soler, M. (ed.), *Evolución: la base de la biología*, Granada, Proyecto Sur de Ediciones, 83-126.

Craver, C. 2006, 'When mechanistic models explain', *Synthese*, 153, 335-376.

Elgin, M. 2006, 'There May Be Strict Empirical Laws in Biology, after All', *Biology and Philosophy*, 21, 119-134.

[10] Como propone Roman Frigg (2006). En nuestro caso, sería sumamente interesante comparar los modelos biológicos matemáticos (como los ya mencionados) con otros, hoy en día matematizados, pero de origen diferente (como el modelo quimiosmótico de Mitchell, cuyo origen es un modelo físico).

Elgin, M. & Sober, E. 2002, 'Cartwright on explanation and idealization", *Erkenntnis*, 57, 441-450.
Fontdevila, A. & Moya, A. 2003, *Evolución: Origen, adaptación y divergencia de las especies*, Madrid, Síntesis.
Frigg, R. 2006, 'Scientific Representation and the Semantic View of Theories', *Theoria* 55, 49-65.
Iranzo, V. 2011, 'Ciencia, modelos, ¿ficciones?', *Teorema*, XXX/2, 157-173.
McMullin, E. 1978, 'Structural Explanation', *American Philosophical Quarterly*, 15 (2), 139-147.
—1985, 'Galilean idealization', *Studies in History and Philosophy of Science*, 16, 247-273.
Ruse, M. 1979, *La filosofía de la biología*, Madrid, Alianza.
Smart, J. J. C. 1963, *Philosophy and Scientific Realism*, London, Routledge & Kegan Paul.
Sober, E. 2000, *Philosophy of biology (Second Edition)*, Oxford, Oxford University Press.

IMPLICACIONES FILOSÓFICAS DE LA BIOLOGÍA POST-GENÓMICA

ALFREDO MARCOS
Universidad de Valladolid (España)
amarcos@fyl.uva.es

ABSTRACT: Here I revise the recent history of Biology since the discovery of the structure and function of DNA, until the completion of the HGP. This journey teaches us something important for philosophy: a successful reductionist approach, focusing on the genetic level, has come to reveal the limits of the ontological reductionism. Not everything is in the genes. So, we should look at the living beings from another point of view. We should not see them merely as vehicles for their genes, but as systems in their own right. This shift of paradigm in biology has interesting philosophical implications. Philosophers should help in defining the concept of biosystem, in deciding which units count as biosystems, and in considering a plural causality, with different directions and levels.

I will argue for the characterization of biosystems by its internal relations, its history, differential development, complexity, functionality and identity. Consequently, the organisms will be presented as the entities that paradigmatically count as biosystems. The parts of organisms are also biosystems, but in a secondary sense, as well as other biological entities of higher level, such as families, populations or ecosystems. On the other hand, abstract entities, such as species, cannot count as biosystems at all. Finally, regarding causation, the new post-genomic perspective invites us to recognize in biology an inhomogeneous plurality of causal directions.
KEY WORDS: Post-genomic biology, Systems Biology, causality, reductionism

1. Del reduccionismo genético a la biología sistémica

"Una revolución está ocurriendo en biología". Con esta frase comienza un artículo publicado en la revista *Genome Research* en el año 2003 (Ehrenberg *et al*. 2003). Casi al mismo tiempo se presentaban los resultados definitivos del Proyecto Genoma Humano. La situación parece paradójica. En el PGH se invirtieron sobre 90.000 millones de dólares y el trabajo de numerosos científicos durante más de una década. Fue el mayor exponente de una forma de hacer biología, según la cual la clave de la vida está en los genes. En consecuencia, si conseguimos secuenciar el genoma humano, tendremos en nuestras manos las claves de la vida humana. Tendremos la posibilidad de identificar los genes que producen enfermedades. Estará a nuestro alcance la prevención, e incluso la predicción. También la curación mediante terapias génicas.

Estas esperanzas se gestaron durante la época de auge de la perspectiva reduccionista. Fijemos el inicio de dicha época en la década de los cincuenta, con el descubrimiento de la estructura del ADN y del funcionamiento del código genético. Pero quizá la imagen más conspicua del reduccionismo genético la hallamos en el famoso libro de Richard Dawkins titulado *El gen*

egoísta (Dawkins 1976). Aquí los organismos desaparecen prácticamente de la ontología biológica. Pasan a ser considerados como meros epifenómenos de los genes, meros vehículos.

En 2003 ya se sabía que el PGH había resultado un éxito, pues había logrado la secuenciación del genoma humano. Pero, por otro lado, el PGH había resultado un fracaso. Al parecer, no todo está en los genes. Para empezar, aparecieron menos genes de los que esperábamos, algo menos de 30.000. Parte del material genético no es significativo. La expresión de los genes está modulada por otros genes y por factores epigenéticos. Los factores epigenéticos son hasta cierto punto heredables. En la expresión genética y en la diferenciación celular influyen los factores ambientales. Muchos genes intervienen en la construcción de un solo rasgo fenotípico. Y hay genes que intervienen en la constitución de muchos rasgos...

Esta es la paradoja. Éxito y fracaso al mismo tiempo. ¿Cómo podemos entenderlo? En realidad, el PGH fue un éxito de la investigación biológica. Sin embargo, no cumplió con todas las esperanzas que se habían depositado en él. De ahí la sensación de fracaso. Pero estas esperanzas estaban basadas en una metodología y en una ontología reduccionista. La sensación de fracaso, pues, no procede de los resultados del propio PGH, sino del desajuste entre estos y las expectativas generadas por la mentalidad reduccionista.

El reduccionismo como método ha producido buenos resultados y grandes cantidades de datos. Sin embargo, el reduccionismos ontológico se ha mostrado finalmente erróneo. No es cierto que todo esté en los genes. No es verdad que los organismos sean meras máquinas de supervivencia manipuladas por los genes. La metáfora del gen egoísta confunde más que orienta. Denis Noble propone pasar a la metáfora del gen cautivo por el organismo (Noble, 2006).

La biología actual nos enseña que, más allá de los genes, el nivel epigenético es muy importante, las proteínas tienen un papel fundamental, así como las rutas metabólicas en las que actúan. El entorno celular y tisular condicionan el desarrollo del organismo. El propio organismo goza de cierta autonomía y es agente de su desarrollo y comportamiento. Y, en general, el medio ambiente tiene que ser tomado en cuenta para entender la vida. Por ello se habla ya de biología post-genómica. Por ello se han puesto en marcha las *omic sciences*.

Pero si cambiamos el foco de este modo, entonces estamos haciendo ya otro tipo de biología. En términos de Thomas Kuhn: ha cambiado el paradigma. Es justo decir que “una revolución está ocurriendo en biología”. Nuestra obsesión no son solo los genes, sino la vida en toda su complejidad y en todos sus niveles, en sus aspectos dinámicos y relacionales, la vida desde una perspectiva no-lineal, holista, sintética y sistémica. Paradójicamente, el éxito del PGH, incluso el éxito de la metodología reduccionista, ha producido el fracaso de la ontología reduccionista. Porque nos ha mostrado sus límites. En la era post-genómica emerge un nuevo paradigma, que nos

invita a mirar la biología desde una perspectiva sistémica. ¿Qué implica este cambio para la filosofía?

2. Una mirada filosófica

O'Malley y Dupré, en su artículo *Fundamental issues in systems biology*, aseguran que en este nuevo marco teórico, la filosofía ha de cumplir con nuevas tareas que podemos concretar en tres preguntas: "¿Qué es un sistema? ¿Qué unidades biológicas cuentan como sistemas? ¿Cómo influyen los sistemas sobre sus componentes?" (O'Malley y Dupré 2005, 1274-5). En la respuesta que demos a estos interrogantes se juega gran parte de las oportunidades y riesgos de la nueva biología post-genómica.

2.1 ¿Qué es un biosistema?

La primera pregunta es de carácter conceptual. El término sistema procede del griego. Quiere decir composición, pero no yuxtaposición. Un sistema está formado por elementos colocados conjuntamente y relacionados entre sí. Está constituido por sus elementos y también por las relaciones que estos mantienen. Presenta propiedades que surgen de estas relaciones y que no se pueden obtener simplemente por yuxtaposición o suma de las propiedades de los elementos. Hay sistemas naturales, como el sistema solar, otros artificiales, como un sistema informático. Pero nuestro centro de interés aquí son los sistemas vivos. A diferencia de los artefactos, los vivientes no se construyen ensamblando piezas preexistentes. De hecho, no constan de piezas, sino de partes. Y estas partes no se ensamblan, sino que se constituyen por diferenciación, al mismo tiempo que se va desarrollando el propio sistema. Dicho de otro modo, los sistemas vivos tienen historia, filogénesis y ontogénesis, evolución y desarrollo. Los sistemas vivos no son simplemente sistemas, son un determinado tipo de sistema. Quizá sería adecuado hablar de "biosistemas", para tenerlo siempre presente.

Ahora bien, no se puede mirar sólo una parte de la cuestión. Es cierto que los biosistemas son dinámicos, pero también lo es que preservan su identidad a lo largo de la vida. En parte es la base genética la que permite esta constancia, pues se mantiene desde el inicio de la vida hasta el final de la misma y se repite en todas las células de los organismos. Es cierto que no todo está en los genes, pero esto no debería llevarnos al extremo contrario. La información biológica no reside en los genes, ni en ninguna otra parte. La información es una relación (Marcos 2011). En esa relación interviene un mensaje genético, un contexto celular y una función. Ahora bien, al elegir los genes como mensaje, nuestra elección no es arbitraria. Se trata de biomoléculas muy específicas y flexibles y cuyos cambios pueden ser críticos para el organismo. Una de las implicaciones que

podemos obtener a partir de la historia reciente de la biología es que se requiere equilibrio: es adecuada la crítica que se hace al reduccionismo ontológico, pero no deberíamos olvidar la especificidad del material genético ni su contribución a la identidad. Es adecuado el énfasis en lo dinámico, pero no deberíamos olvidar los factores de estabilidad e identidad a lo largo de la vida y de las generaciones. Es adecuado el énfasis en las relaciones, pero también son importantes los propios elementos.

Es cierto que los biosistemas son complejos. En ellos, las partes de "vuelven" las unas hacia las otras, se "pliegan". De ahí la complejidad (*cum-plexum*) de los sistemas. Pero los biosistemas poseen una complejidad que es más funcional que estructural. Presentan más organización que simple orden. Nada puede ser calificado como biosistema si no posee una complejidad *con sentido*. Una complejidad orientada. Y el sentido lo da la función. A su vez, la función remite al organismo. Nada es funcional si no es en relación con un organismo concreto. Calificamos como funcional la división celular que se produce normalmente dentro de un tejido. Pero calificamos como disfuncional el mismo proceso cuando se da en un tumor. Obviamente la funcionalidad se establece por relación a un ser vivo concreto. Así pues, la biología post-genómica, y en especial su enfoque sistémico, abre necesariamente cuestiones de carácter ontológico sobre la naturaleza de los propios sistemas. Los sistemas que estudiamos podrían resultar meras agrupaciones convencionales, o bien genuinas sustancias dotadas de realidad objetiva. Su complejidad y funcionalidad pueden depender en última instancia de nuestra perspectiva (tal como lo interpreta, por ejemplo, Cummins en su teoría de las funciones), o bien de la realidad de las cosas. Estas observaciones nos conducen ya a la segunda pregunta.

2.2 ¿Qué unidades biológicas cuentan como biosistemas?

Esta segunda cuestión es de carácter empírico. Se trata de ver qué unidades biológicas cumplen con las condiciones que hemos atribuido a un biosistema. Un animal, una planta, un hongo, una bacteria, una persona… Cada organismo vivo es paradigmáticamente un biosistema. Es el punto de referencia principal de la funcionalidad y de la complejidad. Sin embargo, la vida, como es sabido, se estructura jerárquicamente. Otros niveles de lo vivo también podrían contar como biosistemas. En un sentido primario lo son los organismos, pero en un sentido secundario son biosistemas también las partes de los mismos: sistemas, como el digestivo, órganos, como el corazón, tejidos, como el epitelial, células integradas en metazoos y metafitas. En un sentido aun más degradado, contarían como biosistemas también entidades como los simbiontes, los orgánulos celulares o los virus.

Es difícil saber en qué punto de esta escala habría que colocar las entidades que están en niveles superiores al organismo, como las poblaciones, colonias, familias, clanes, otras entidades sociales, y, en definitiva, los ecosistemas. Pero, desde luego, no son biosistemas en un sentido tan neto, primario y paradigmático como lo son los organismos. El caso de las especies y otras unidades taxonómicas también ofrece dificultades. No parece que puedan ser vistos como biosistemas, sino más bien como entidades abstractas, conceptos, con base real (Marcos, 2008).

Todo lo dicho implica un trabajo filosófico. Se trata de construir una ontología que distinga claramente lo abstracto, como los conceptos, de lo concreto, como los biosistemas. Además, dentro de estos deberíamos hacer distinciones de grado. Deberíamos distinguir un sentido primario y paradigmático que sólo sería atribuible a los organismos, y un sentido secundario o derivado en el que estarían otras unidades biológicas, tanto inferiores como superiores a los organismos.

Aquí reside uno de los riesgos de la biología post-genómica que ha de ser identificado y tratado desde el punto de vista filosófico. Podríamos estar saliendo de un reduccionismo demasiado radical para entrar en un holismo demasiado radical. Según la metáfora del gen egoísta, es el gen la auténtica sustancia. El organismo queda *desustancializado*. Y con ello *desvalorizado*. Pero quizá ahora nos dirigimos hacia el ecosistema como auténtica sustancia, con un resultado similar por lo que a la importancia del organismo se refiere. ¿Cómo evitarlo?, ¿cómo reconocer a los ecosistemas y entidades sociales su peso ontológico sin degradar los organismos?

Se podría pensar que en realidad los niveles más importantes son los superiores. Nos guiamos aquí por una imagen espacial, lo pequeño forma parte de lo grande, las células de los tejidos, y los organismos de los ecosistemas. Los organismos están formados por moléculas, células, tejidos, órganos y miembros, algunos agrupados en sistemas. Pero también están formados por elementos sociales, poblacionales y ecosistémicos. Digámoslo así: cada uno de los organismos está hecho de células, pero también de entidades sociales y ecosistémicas. Esta visión permite dar la importancia adecuada a las entidades más amplias, como los ecosistemas, sin anular o instrumentalizar por ello los organismos vivos.

2.3 ¿Cómo influyen los biosistemas sobre sus componentes?

Esta tercera cuestión tiene que ver con la causalidad. Se ha consolidado la distinción entre tres tipos de causalidad en sistemas vivos: *bottom-up*, *top-down* y *middle-out* (Noble 2006). La primera dice que los cambios en las partes causan efectos en el sistema como un todo. La segunda que el sistema como tal es capaz de causar efectos sobre sus partes. La tercera diría que podemos partir de cualquier nivel de la jerarquía biológica y encontraremos que la actividad en dicho nivel produce efectos en los niveles contiguos, tanto hacia arriba como hacia abajo.

La biología post-genómica ha de reconocer todas estas direcciones de la causalidad, pero no hay simetría entre la causalidad *bottom-up* y la causalidad *top-down*. Contrariamente a lo que se suele pensar, la segunda es más empírica, la primera más teórica y conjetural. Tenemos experiencia directa de causalidad *top-down*, dado que nosotros, como organismos que somos, producimos efectos mediante nuestra libre acción sobre las partes. Por ejemplo, podemos cambiar la postura de nuestro cuerpo, la flexión de un brazo o el contenido en colesterol de nuestra sangre. En el sentido contrario, *bottom-up*, sólo registramos correlaciones, sobre las que proyectamos, mediante conjeturas teóricas, una relación causal. En realidad, no podemos tener experiencia directa de causalidad *bottom-up*, como anticipó Hume. Por ejemplo, tras un cierto cambio hormonal sentimos estrés o tranquilidad. Conocemos la secuencia temporal y la correlación, pero no tenemos experiencia de la causalidad, pues al fin y al cabo, la tranquilidad, o el estrés en tanto que sensaciones, son heterogéneos respecto de las hormonas o de los neurotransmisores en tanto que moléculas. Si hay causalidad empírica, esa es la causalidad *top-down*. Con esto no se quiere decir que no haya causalidad *bottom-up*, sino que se trata de un tipo de causalidad que conocemos de modo más indirecto, teórico y conjetural. Una biología centrada en los sistemas más que en los genes es una ciencia que ha de tener en cuenta todas las direcciones de la causalidad, pero también ha de saber cuál es primaria y cuál derivada.

Por lo que hace al concepto de causalidad *middle-out*, introducido por Noble, constatamos que podría introducir un sesgo relativista, evitable con un cuidadoso análisis filosófico. Es cierto que todos los niveles de la jerarquía biológica tienen capacidad causal y ejercen efectos sobre otros adyacentes. Pero no todos tienen la misma importancia causal. Son los organismos los que prioritariamente han de ser considerados como biosistemas. Se puede conjeturar, pues, que serán también los agentes causales más relevantes para la explicación biológica, tanto de los niveles inferiores como de los superiores. La oportunidad que se abre ante nosotros con la nueva biología se hará más real si desarrollamos una ontología que podríamos denominar *pluralismo no homogéneo*. Esto es tanto como reconocer realidad ontológica y capacidad causal a todos los niveles de la vida, pero muy especialmente a los organismos.

3. Conclusión

Hemos atendido a la historia reciente de la biología, desde el descubrimiento de la estructura y función del ADN, hasta la culminación del PGH. Este recorrido nos ha enseñado algo importante para la filosofía. Una exitosa metodología reduccionista, centrada en el nivel genético, ha llegado a poner de manifiesto los límites del reduccionismo ontológico. No todo está en los genes. A partir de ahí hemos comenzado a mirar de otra forma a los seres vivos. No los vemos ya como meros

vehículos de sus genes, sino como sistemas con entidad propia. Este cambio de paradigma en biología tiene implicaciones filosóficas. Los filósofos han de pensar *i*) el concepto mismo de biosistema. *ii*) Deben contribuir a decidir qué unidades cuentan como biosistemas. Y, además, *iii*) deben considerar el problema de una causalidad plural, con diferentes direcciones y niveles.

i) Se ha argumentado a favor de la caracterización de los biosistemas por sus relaciones internas, su historicidad, desarrollo diferencial, complejidad, funcionalidad e identidad. *iii*) En consecuencia, las entidades que cuentan paradigmáticamente como biosistemas son los organismos. También son biosistemas las partes de dichos organismos y otras entidades biológicas concretas de nivel superior, como por ejemplo las familias, poblaciones o ecosistemas. No cuentan, en cambio, como tales biosistemas las entidades abstractas, como pueden ser las especies. *iii*) Por último, y en cuanto a la causalidad, la nueva perspectiva post-genómica nos invita a reconocer en biología una pluralidad no homogénea de direcciones causales, no solo ascendentes, sino también descendentes.

REFERENCIAS

Dawkins, R. 1976, *The Selfish Gene*, Oxford, Oxford University Press.

Ehrenberg, M. *et al*. 2003*,* 'Systems Biology is Taking Off', *Genome Research* 13, 2377-2380.

Marcos, A. 2011, 'Bioinformation as a Triadic Relation', en Terzis, G. y Arp, R. (eds.), *Information and Living Systems: Philosophical and Scientific Perspectives*, Cambridge, Massachusetts, M.I.T. Press, 55-90.

Marcos, A. 2008, 'The Species Concept in Evolutionary Biology: Current Polemics', en González, W. J. (ed.), *Evolutionism: present approaches*, La Coruña, Netbiblo, 121-142.

Noble, D. 2006, *The Music of Life. Biology Beyond the Genome*, Oxford, Oxford University Press.

O'Malley, M. A. y Dupré, J. 2005. 'Fundamental issues in systems biology', *BioEssays* 27, 1270–1276.

MEDIDAS PROBABILISTAS DE RELACIÓN EVIDENCIAL: ¿PLURALIDAD DE MEDIDAS O PLURALIDAD DE RELACIONES?

IGNACIO MARTÍNEZ DE LEJARZA Y ESPARDUCER
Universidad de Valencia.
e-mail: mlejarza@uv.es

ABSTRACT: This paper argues that the various (and not equivalent) probabilistic measures of confirmation correspond to different sorts or conceptual aspects of the evidential support relation. It seems necessary to distinguish between confirmation and explanation, on the one hand, and between gross and net evidential support, on the other hand. Finally considering some different issues involved for each measure can establish an appropriate application scope.
KEYWORDS: Confirmation, explanation, confirmation measures, evidential relation, explanation power measures, Bayesian Epistemology.

Introducción

Dar cuenta de las relaciones evidenciales en términos probabilísticos es una opción recurrente en Filosofía de la Ciencia. En particular, el traducir a términos probabilísticos la confirmación de hipótesis a través de la experiencia ha sido una cuestión defendida desde diversas posiciones y particularmente desde el bayesianismo. Recientemente también se han propuesto (Schupbach y Sprenger (2011); Tentori y Crupi (2012)) distintas aproximaciones probabilistas para medir el grado de explicación que una determinada hipótesis o teoría le confiere a cierto resultado experimental o empírico.

Sin embargo, no parece haber un claro consenso en la manera particular de cuantificar probabilísticamente estas relaciones evidenciales. En este trabajo intentaremos analizar cómo estas diferentes aproximaciones pueden corresponderse legítima y también diferenciadamente con distintos matices de la relaciones evidenciales. Este enfoque iniciado por Joyce (2004) y (2005), conlleva la discusión y clarificación de un adecuado mapa conceptual de las diferentes relaciones evidenciales que puedan establecerse. El objetivo sería aclarar, no solo las diferencias numérico-cuantitativas, sino también las condiciones y ámbito de aplicación de las medidas propuestas.

1. La representación probabilista de las relaciones evidenciales

Desde el bayesianismo las relaciones evidenciales pueden ser óptimamente representadas de forma probabilística siempre y cuando asumamos, como señala Joyce (2005), que la asignación

de probabilidades en un momento *t* y condicionada a toda la información (empírica y no empírica) conocida, debe recoger toda la evidencia disponible (evidencia total disponible en *t*). Añadiendo también un sentido de marcado carácter "logicista" de forma que la asignación de probabilidades sea tal que para cualquier proposición para la que la probabilidad esté definida ésta deba reflejar el "balance total de razones en favor de la verdad de H".

Sobre este esquema es posible concebir dos formas diferentes de representar la evidencia disponible a favor de cierta proposición X en un momento dado y con el conocimiento disponible en ese momento. Bien de una forma directa a través de la asignación de probabilidades P, como P(X); o bien estableciendo una medida de apoyo empírico que nos considere la evidencia a favor de la veracidad de X frente a la evidencia a favor de su falsedad; esto es P(X)/P(¬X). En términos de Joyce (2004) hablaríamos respectivamente de medida de la *evidencia bruta* y de la *evidencia neta* en favor de X.

Esta segunda medida se corresponde con la función que a menudo se designa como *Odds*, palabra inglesa que viene a significar la proporción a la que se está dispuesto a pagar una apuesta en favor de algo y que refleja el cociente entre la creencia a favor y en contra de su ocurrencia.

Como la relación funcional entre probabilidades y *Odds* no es lineal, como puede verse en la figura 1, los posibles cambios de ambas representaciones de la evidencia en función del registro de nueva evidencia (vía condicionalización probabilística) no obedecerán en general al mismo patrón.

Figura 1: relación entre Odds y probabilidad

De esta forma, si al obtener una nueva evidencia, la probabilidad de la hipótesis *h* se incrementa (decrementa), hablaremos de incremento (decremento) bruto de la evidencia a su favor y si es la función Odds la que se incrementa (decrementa), hablaremos de incremento (decremento) neto de la evidencia a favor.

Es precisamente esta distinción la que utilizará el propio Joyce para desarrollar su idea de una pluralidad de sentidos diferentes de confirmación que pueden, y deben, no ser equivalentes. Veamos esta cuestión con un poco más de detalle.

2. Las distintas medidas (no equivalentes) de confirmación y la disolución del problema.

Una de las ventajas de un enfoque probabilista de la confirmación es obviamente la posibilidad de cuantificar el apoyo empírico que recibe una hipótesis al incorporarse una nueva evidencia. Sin embargo, más que contar con una única medida de apoyo evidencial, distintas y variadas medidas parecen dar cuenta de la confirmación (incremental) que recibe una hipótesis al incorporarse una nueva experiencia. Lo problemático de esta situación es el conocido problema de la "*no equivalencia ordinal*" entre ellas (Fitelson, 1999 ,2001); no sólo dan distintos valores numéricos, lo que podría no ser problemático, sino que pueden llegar a establecer ordenes diferentes al evaluar el apoyo que distintas evidencias prestan a distintas hipótesis.

Este problema se ha intentado resolver por dos vías distintas: defendiendo una medida de confirmación evidencial particular (postura mayoritaria, aunque con opciones divididas), o disolviendo el problema concluyendo que no hay una forma precisa de medir la "confirmabilidad".
Las medidas particulares de confirmación mayoritariamente defendidas se basan todas ellas en una suerte de comparación o bien entre la probabilidad de la hipótesis h antes ($P(h)$) y después ($P(h|e)$) de la incorporación de la evidencia e, o bien entre la razón de probabilidades a favor y en contra de h , también, antes ($P(h) /P(\neg h)$) y después ($P(h|e) /P(\neg h|e)$) de incorporar la evidencia e. Lo que ya más arriba hemos designado como Odds, de forma que se trataría de comparar $Odds(H)$ y $Odds(H|e)$.

Autores como Jeffrey (1992) , Earman (1992), Rosenkratz (1977) , Howson y Urbach (2006) han defendido, como medida de confirmación, la diferencia entre las probabilidades. Keynes (1921), Horwich (1982) , Schlesinger (1995), Milne (1995) y (1996), Kuipers (2000) han defendido el cociente entre las probabilidades, normalizándolo, con su logaritmo, o no . En cambio, Good (1984), Howson (2003), Jeffrey (2004) y Fitelson (2006) han propuesto el cociente entre las verosimilitudes de h y de $\neg h$, cociente equivalente al cociente entre $Odds(H)$ y $Odds(H|e)$.
Las razones aducidas para la defensa de unas u otras medidas suelen obedecer al cumplimento o no de ciertas condiciones o a la presentación de situaciones en las que se apela a la intuición para discernir la mejor medida. Sin embargo, que existan distintas medidas del "grado en que una nueva evidencia confirma (o disconfirma) una hipótesis" bien puede obedecer al hecho de que cada

medida da cuenta de una particularidad diferente de la confirmabilidad incremental. Esta línea ha sido defendida por James Joyce, y por Alan Hájek (2008), entre otros.

La distinción de Joyce (2004) , de la que ya hemos hablado, entre evidencia total *bruta* en la que la relación evidencial se mide según un criterio probabilista absoluto (utilizando probabilidades y midiendo su variación), y evidencia total neta, en la que la relación evidencial se mide según un criterio probabilista relativo (utilizando *odds*, cocientes entre probabilidades a favor y en contra, y su variación) puede sernos muy útil para empezar a remarcar un mapa conceptual de quizá muy diferentes tipos de relaciones evidenciales.

Aunque la propuesta de Joyce que aquí se analiza y se pretende desarrollar está fundamentalmente orientada hacia las medidas de confirmación, es plausible que , también pudiera aportar cierta luz a la hora de abordar la medición de la relación de "explicación de un resultado evidencial por cierta hipótesis".

3. Primera propuesta de clasificación

Las relaciones de apoyo empírico o evidencial existentes entre hipótesis explicativas y proposiciones evidenciales parece que pueden clasificarse, atendiendo, por un lado, a la dirección u orden de la relación de forma que encontramos un primer factor (A) que permitiría distinguir entre:
(A1, orden directo) el *apoyo* que la evidencia confiere a la hipótesis y
(A2, orden inverso) el mejor o peor *poder* "*explicativo*/*predictivo*" que la hipótesis confiere a la evidencia.
Por otro lado, atendiendo a un factor de alcance o de aplicación (B) cabría distinguir entre:

(B1, de alcance absoluto) la relación evidencial (ya sea de confirmación o de explicación) entendida como una mejora (o merma) en la acumulación de evidencia en favor de una hipótesis o, simétricamente, como una mejora (o merma) en la capacidad de la hipótesis para explicar el acervo evidencial y

(B2, de alcance comparativo) la confirmación entendida como incremento en la credibilidad de una hipótesis en detrimento de su negación o de la hipótesis rival, cuestión que apunta a situaciones diagnóstico o *contrastivas* en un amplio sentido. O, alternativamente, el incremento en el poder explicativo que la introducción de la hipótesis en cuestión confiere a una evidencia, en detrimento del poder explicativo que su introducción aportaría a la evidencia contraria.

La siguiente tabla reflejaría la doble clasificación:

		Dirección u Orden (A)	
		Directo (A1)	Inverso (A2)
Alcance (B)	Absoluto (B1)	*(A1,B1) Confirmación de alcance absoluto*	*(A2,B1) Explicación/predicción de alcance absoluto*
	Comparativo (B2)	*(A1,B2) Confirmación de alcance comparativo*	*(A2,B2) Explicación/ predicción de alcance comparativo*

Tabla 1. Clasificación de las relaciones evidenciales según dirección y alcance

Como argumentaremos más tarde, para el caso de la confirmación (columna A1), parece que la situación A1B1 podría medirse adecuadamente con la razón de probabilidades inicial y final y la situación A1B2 con la razón de verosimilitudes de la hipótesis y su negación;

$$RP(h,e) = \frac{P(h\,|\,e)}{P(h)} \quad , \quad RL(h,e) = \frac{Odds(h\,|\,e)}{Odds(h)} = \frac{P(e\,|\,h)}{P(e\,|\,\neg h)}$$ respectivamente.

Sin embargo, para el caso de la explicación (columna A2) el asunto no acaba de resultar de medición evidente.

Así, una medida de la explicación (de alcance absoluto) de la evidencia e por la hipótesis h basada en la razón de probabilidades resultaría idéntica a la propuesta para la confirmación absoluta.

$$\boxed{MEP(e,h) = \frac{P(e\,|\,h)}{P(e)} = \frac{P(e\,|\,h).P(h)}{P(e).P(h)} = \frac{P(h\,|\,e)}{P(h)} = RP(h,e)}$$

Ecuación 1

Y una medida de la explicación de la evidencia e por la hipótesis h de alcance relativo basada en las “Odds” acaba resultando equivalente a expresiones de difícil interpretación.

$$MEO(e,h) = \frac{Odds(e\,|\,h)}{Odds(e)} = \frac{\dfrac{P(e\,|\,h)}{P(\neg e\,|\,h)}}{\dfrac{P(e)}{P(\neg e)}} = \frac{\dfrac{P(e\,|\,h)}{P(e)}}{\dfrac{P(\neg e\,|\,h)}{P(\neg e)}} = \frac{\dfrac{P(h\,|\,e).P(e)}{P(e).P(h)}}{\dfrac{P(h\,|\,\neg e).P(\neg e)}{P(\neg e).P(h)}} = \frac{P(h\,|\,e)}{P(h\,|\,\neg e)}$$

Ecuación 2

Por tanto, la relación de explicación o predicción, al no ser meramente una inversión de la relación de confirmación, no parece que pueda cuantificarse en formato probabilístico por medio de una simple inversión de los argumentos de las funciones.

4. Medidas brutas y netas de confirmación

Vamos a desarrollar en este epígrafe las medidas que pretenden dar cuenta de las relaciones evidenciales de orden directo (Columna A1 de la Tabla 1) tanto para su alcance absoluto como para su alcance neto o contrastivo.

Esta parte de nuestra clasificación se corresponde con la distinción establecida por Joyce entre evidencia total bruta y evidencia total neta:

Si damos por hecho que la evidencia total bruta disponible en favor de una hipótesis *h* en un momento t y ante un cuerpo de conocimiento dado queda registrada por su probabilidad, *P(h)* y que la evidencia neta en favor de *h* en relación a la evidencia en su contra queda registrada por *Odds(H)*, tendremos que :

Una nueva evidencia *e*:

A) confirmará/ resultará irrelevante /disconfirmará (incrementalmente) *de forma bruta*
si y sólo si *P(h |e) > /=/ < P(h)*
mientras que :

B) confirmará/ resultará irrelevante /disconfirmará (incrementalmente) *de forma neta*
si y sólo si *Odds(h|e) > /=/ < Odds(h)*

A partir de aquí, para contar con un conjunto de medidas del *grado* de confirmación incremental de cada tipo nos basta con introducir una a varias funciones de comparación. Las más habituales serán la diferencia y el cociente entre la evidencia (bruta o neta) antes y después de la ocurrencia de *e* :

Medidas de incremento evidencial	Bruto	Neto
Razón	*P(h|e) / P(h)*	*Odds(h|e)/Odds(h)*
Diferencia	*P(h|e) – P(h)*	*Odds(h|e) – Odds(h)*

Tabla 2. Medidas de incremento evidencial (confirmación)

De entre las cuatro medidas propuestas en la tabla de arriba, suelen preferirse las medidas de razón por una mayoría de autores entre los que podemos citar: Fitelson (2006), Iranzo y Mtz. de Lejarza (2010), Good (1983), Horwich (1982), Howson (2003), Jeffrey (2004), Keynes (1921), Kuipers (2000), Milne (1995) y (1996), Schlesinger (1995), Zalabardo (2009). En buena parte de estos trabajos se suele criticar el hecho de que las medidas basadas en la diferencias suelen arrastrar una clara dependencia de las probabilidad a priori.

Merece la pena detenerse un poco sobre este particular así como desarrollar algunas equivalencias que interesa considerar.

En primer lugar a las medidas de razón siguiendo a Zalabardo (2009) e Iranzo y Mtz. de Lejarza (2012) las llamaremos PR : *probability ratio* y LR: *likelihood ratio*. La razón de la primera es obvia y la de la segunda puede comprobarse muy fácilmente:

$$PR(h,e) = \frac{P(h\,|\,e)}{P(h)} = \frac{\dfrac{P(e\,|\,h).P(h)}{P(e)}}{P(h)} = \frac{P(e\,|\,h)}{P(e)}$$

$$\mathrm{LR(h,e)} = \frac{\mathrm{Odds(h|e)}}{\mathrm{Odds(h)}} = \frac{\dfrac{P(h\,|\,e)}{P(\neg h\,|\,e)}}{\dfrac{P(h)}{P(\neg h)}} = \frac{\dfrac{P(e\,|\,h)}{P(e)}}{\dfrac{P(e\,|\,\neg h)}{P(e)}} = \frac{P(e\,|\,h)}{P(e\,|\,\neg h)}$$

Ecuaciones 3

De estos sencillos desarrollos se desprende, por un lado, que ni PR ni LR dependen de las probabilidades a priori o dicho de otra forma de la acumulación previa de evidencia en favor de *h* y sólo consideran el incremento de esa evidencia. Por otro lado, es fácil observar cómo la medida LR da cuenta de la variación bruta del apoyo en favor de *h* al disponer de la evidencia *e* , dividida por la variación bruta del apoyo en contra de *h* (a favor de ¬*h*): Formalmente: *LR(h,e)=PR(h,e)/PR(¬h,e)*. Como cabría esperar LR supone la consideración de la confirmabilidad neta: balance de la confirmación a favor y en contra, mientras que PR supone sólo la consideración de la confirmabilidad bruta (la variación del apoyo en favor de *h*)

Ambas medidas captan adecuadamente aspectos bien distintos de la variación en el apoyo empírico que supone una nueva evidencia: Mientras en PR simplemente nos daría información sobre la mejora o empeoramiento de la evidencia en favor de cierta hipótesis, LR nos daría información sobre la mejora o empeoramiento de la plausibilidad de la hipótesis frente a su contraria. Esto sitúa a PR en un contexto de medición que, a mi juicio sería más estrictamente confirmativo, mientras que LR quedaría situada en un contexto más contrastivo, diagnóstico, o decisional.

En los siguientes gráficos podemos intuir el distinto comportamiento de ambas medidas en estos dos ámbitos:

En el primero de ellos (Figura 2) se representan en el llamado diagrama Bernardo-Smith (2000) para distintos valores de las mediciones LR y PR sobre el plano coordenado "*probabilidad inicial – probabilidad final*", o sea *P(h)* y *P(h|e)*. Se observa cómo PR registra más claramente y directamente la variación de la probabilidad antes y después de observase *e* que lo que lo hace LR. Una PR constante nos habla de situaciones en las que se ha transitado de *P(h)* a *P(h|e)* en la misma proporción mientras que una LR constante, no.

Figura 2. Diagrama Bernardo-Smith para LR y PR

En cambio en la Figura 3, el llamado diagrama ROC, Receiver Operating Characteristic, (Tanner y Swets, 1954),(Swets,1995), nos da cuenta del balance entre sensibilidad e inespecificidad que se da para cada situación de *PR= constante* y *LR= constante*. Mientras que para todos aquellos casos en los que LR tome un mismo valor constante el balance es idéntico y no depende de la probabilidad inicial de *h*; no ocurre igual para PR ya que, aunque dos situaciones puedan compartir el mismo valor de PR, nada garantiza que supongan el mismo balance sensibilidad/efectividad y de hecho, este balance mejorará para dos situaciones de idéntico PR si la probabilidad inicial es mayor como ocurre en la figura 3 al pasar de *P(h)* =0.316, a la izquierda, a *P(h)* =0.6 ,a la derecha.

Por sensibilidad de una prueba dicotómica que puede dar como resultado *e* o ¬*e* se entiende la probabilidad de que siendo cierta la hipótesis, la prueba dé *positiva* es decir *e*, esto es, la probabilidad de un correcto positivo, en términos decisionales, o la verosimilitud de la hipótesis *h*, en términos de inferencia estadística.($P(e|h)$)

Por especificidad se entiende, en cambio, la probabilidad de que la prueba dé negativa, ¬*e*, cuando no se cumpla la hipótesis, esto es la probabilidad de un correcto negativo o la verosimilitud de la hipótesis ¬*h*, cuando la prueba dé negativa ($P(\neg e|\neg h)$). Obviamente la inespecificidad es el valor complementario de la especificidad.

Una prueba será tanto más *exacta* cuanto mayores sean su sensibilidad y su especificidad. Y por tanto, cuanto más nos acerquemos al extremo superior izquierdo en el esquema ROC (punto (0,1)), tendremos una prueba más *exacta*.

Obviamente el mejor comportamiento de LR en el diagrama ROC nos habla de su mejor aplicación en situaciones contrastivas o diagnósticas.

Figura 3 Diagrama ROC para LR=cte y PR=cte

Por lo que hace a las medidas basadas en diferencias es sencillo probar que en ambos casos existe una relación lineal directa de cada una de ellas con su correspondiente medida de razón, en la que el factor de proporcionalidad depende linealmente y en sentido directo, o positivo, de la evidencia previa (probabilidad inicial antes de *e*; o bien, Odds inicial, antes de *e*):

$$Dp(h,e) = P(h \mid e) - P(h) = \left(\frac{P(h \mid e)}{p(h)} - 1 \right) P(h) = \left(PR(h,e) - 1 \right).P(h)$$

$$Do(h,e) = Odds(h \mid e) - Odds(h) = \left(\frac{Odds(h \mid e)}{Odds(h)} - 1 \right) Odds(h) = (LR(h,e) - 1).Odds(h)$$

Esta dependencia de las probabilidades iniciales hace, a mi juicio, menos interesantes estas medidas en un contexto estrictamente confirmatorio, sin embargo, en situaciones prácticas en las que se debe optar por mantener o descartar una cierta hipótesis y donde el contexto nos exige no despreciar la información inicial pueden ser medidas de gran interés. Un claro ejemplo de este tipo de situaciones puede ser el diagnóstico médico en el que consideraciones sobre la prevalencia (presencia general o a priori) de cierta enfermedad no deben despreciarse a la hora de valorar una prueba diagnóstica.

5. Las medidas de explicación:

Como ya hemos comentado, las medidas de explicación no acaban de encajar en un esquema de desarrollo en el que, siguiendo con la idea de la tabla 1, la explicación con la que *h* da cuenta de e, en términos absolutos, debería medirse a través del cociente entre *P(e|h)* y *P(e)* y , análogamente, la explicación neta con la que *h* da cuenta de e, en relación a su capacidad para explicar ¬*e* debería medirse a través del cociente entre *Odds(e|h)* y *Odds(e).*

Ya vimos, en las ecuaciones 1 y 2 cómo estos dos cocientes se correspondían respectivamente con la misma medida de confirmación bruta y con la expresión *P(h|e)/P(h|¬e)* que se corresponde con lo que Joyce (2004) llama incremento de evidencia *efectiva* (o confirmación *efectiva*) de *h* por *e*. Una ratio entre el apoyo que *e* le confiere a *h* y el que le confiere ¬*e*.

Obviamente, ninguna de estas dos medidas parece corresponderse con lo que habitualmente se concibe como explicación de un resultado empírico por parte de una hipótesis o teoría.

No parece, por lo tanto, que, al menos de momento, el camino seguido hasta aquí vaya a ser fructífero por lo que se refiere a la relación de explicación. Las medidas que han sido recientemente propuestas, de hecho, no siguen este guión sino que directamente intentan abordar una medida probabilista de la explicación de *e* por *h* partiendo del sentido más habitual de explicación:

Así, por ejemplo, Schupbach y Sprenger (2011) parten de la idea de que el poder explicativo del explanans (la hipótesis *h*) tiene que relacionarse con su capacidad de hacer disminuir en el grado en el que encontremos sorprendente el explanandum (la evidencia *e*). Este sentido tendría

también una medida que comparara la probabilidad (o las Odds) de *e* condicionada(s) a *h* con la probabilidad (o las Odds) incondicional(es). Aunque, como hemos visto, estas medidas no parecen haber resultado fructíferas.

Partiendo de esta idea , Schupbach y Sprenger proponen una serie de condiciones que una adecuada medida del poder explicativo debería reunir , llegando a la conclusión de que la medida adecuada debería ser:

$$E_{ss}(e,h) \underset{def}{=} \frac{P(h\,|\,e)-P(h\,|\,\neg e)}{P(h\,|\,e)+P(h\,|\,\neg e)}$$

Ecuación 4

Por su parte, Crupi y Tentori (2012) sobre un análisis crítico del trabajo de Schupbach y Sprenger proponen , en cambio, una medida de distancia relativa, entre la probabilidad de *e* condicionada a *h* y la probabilidad incondicionada de *e*:

$$E_{CT}(e,h) \underset{def}{=} \begin{cases} \dfrac{P(e\,|\,h)-P(e)}{1-P(e)} & \text{si P(e|h)} \geq \text{P(e)} \\ \dfrac{P(e\,|\,h)-P(e)}{P(e)} & \text{si P(e|h)} < \text{P(e)} \end{cases}$$

Ecuación 5

A pesar de que parece tener todo el sentido dentro del marco general que hemos expuesto hasta ahora, si se desarrolla un poco la medida propuesta por Crupi y Tentori también se puede apreciar una relación de cuasi-coincidencia con una medida de confirmación incremental bruta.
Como una de las formas de expresar la medida de confirmación incremental bruta (razón de probabilidades final/inicial) es :

$PR(h,e) = \dfrac{P(e\,|\,h)}{P(e)}$ podemos manipulando la expresión de la Ecuación 5, obtener:

$$E_{CT}(e,h) \underset{def}{=} \begin{cases} \text{si P(e|h)} \geq \text{P(e):} & \dfrac{\dfrac{P(e\,|\,h)}{P(e)}-1}{\dfrac{1}{P(e)}-1} \\ \text{si P(e|h)} < \text{P(e):} & \dfrac{P(e\,|\,h)}{P(e)}-1 \end{cases} = \begin{cases} \text{si P(e|h)} \geq \text{P(e):} & \dfrac{PR(h,e)-1}{\dfrac{1}{P(e)}-1} \\ \text{si P(e|h)} < \text{P(e):} & PR(h,e)-1 \end{cases}$$

Ecuación 6

Lo que supone que la medida de Crupi-Tentori es una transformación lineal por tramos de la medida de confirmación incremental bruta, PR, como puede verse en el gráfico siguiente, por lo que, más allá de las diferencias en escala y origen, ambas medidas estarían dando cuenta de un fenómeno idéntico.

Figura 4: ECT, y PR en función de la verosimilitud para *P(e)* =0.2 y *P(e)* =0.6

Por lo que hace a la medida de Sprenger –Schupbach que venía dada por:

$$\boxed{E_{ss}(e,h) \underset{def}{=} \frac{P(h\,|\,e)-P(h\,|\,\neg e)}{P(h\,|\,e)+P(h\,|\,\neg e)}}$$

Aplicando el teorema de Bayes y algunas transformaciones sencillas, tendremos que:

$$\begin{aligned}
&E_{ss}(e,h) \underset{def}{=} \frac{P(h\,|\,e)-P(h\,|\,\neg e)}{P(h\,|\,e)+P(h\,|\,\neg e)} = 1-2\frac{P(h\,|\,\neg e)}{P(h\,|\,e)+P(h\,|\,\neg e)} = 1-\frac{2}{\frac{P(h\,|\,e)}{P(h\,|\,\neg e)}+1} = \\
&= 1-\frac{2}{\frac{P(e\,|\,h)P(h)P(\neg e)}{P(e)P(\neg e\,|\,h)P(h)}+1} = 1-\frac{2}{LR(h,e)\frac{(1-P(e))}{P(e)}+1} = \\
&= \frac{LR(h,e)\frac{(1-P(e))}{P(e)}+1-2}{LR(h,e)\frac{(1-P(e))}{P(e)}+1} = \frac{\frac{LR(h,e)}{Odd(e)}-1}{\frac{LR(h,e)}{Odd(e)}+1} = \frac{LR(h,e)-Odd(e)}{LR(h,e)+Odd(e)}
\end{aligned}$$

Ecuación 7

Resultando que ESS es una función desaceleradamente creciente con respecto a la medida de confirmación LR y aceleradamente decreciente con respecto a la probabilidad incondicionada de la evidencia (su factibilidad inicial), *P(e)* .

Figura 5. Ess(e,h) en función de LR(h ,e) para valores de *P(e)* de 0.2, 0.4 y 0.6

6. A modo de (in)conclusión

Por lo que hemos podido ver, parece que el conocido problema de la no equivalencia entre distintas medidas probabilistas de confirmación incremental puede disolverse, al menos parcialmente, aclarando lo que cada una de ellas termina por medir y dando a cada una su adecuada aplicabilidad. El camino sin embargo no parece haber concluido, ni mucho menos, y prueba de ello es que unas recíprocas medidas de explicabilidad aún no han sido establecidas y las propuestas hasta ahora quizá deban todavía ser revisadas. La conclusión a introducir en este foro es la necesidad de un adecuado mapa conceptual de las distintas relaciones evidenciales que nos aclararán no solo las diferencias numérico-cuantitativas, sino también sus condiciones y ámbito de aplicación.

Agradecimientos: Este trabajo ha sido posible gracias a la ayuda del Ministerio de Ciencia e Innovación a través de los fondos destinados al proyecto FFI2008-01169

REFERENCIAS

Bernardo, J.M y Smith, A.F.M. 2000, *Bayesian Theory*, New York, John Willey

Crupi, V., Tentory, K. y González, M. 2007, 'On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues', *Philosophy of Science*, vol. 74, 229-252.

Crupi, V. and Tentori, K. 2012(pendiente de publicación), 'A second look at the logic of explanatory power (with two novel representation theorems)', *Philosophy of Science*.

Earman, J. 1992, *Bayes or Bust. A Critical Examination of Bayesian Confirmation Theory*, Cambridge, MIT Press.

Eells, E. y Fitelson, B. 2000, 'Measuring Confirmation and Evidence', *Journal of Philosophy*, vol. 97, 663-672.

Fitelson, B. 1999, 'The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity', *Philosophy of Science*, vol. 66, S362-S378.

Fitelson, B. 2001, *Studies in Bayesian Confirmation Theory*. Tesis doctoral no publicada. Universidad de Wisconsin. Madison

Fitelson, B. 2002, 'Putting the Irrelevance Back into the Problem of Irrelevant Conjunction', *Philosophy of Science*, vol. 69, 611-622.

Good, I.J. 1984, *Good Thinking. The Foundations of Probability and Its Applications,* Minneapolis, University of Minnesota Press.

Hájek A. & Joyce,J. 2008, 'Con firmation', in Psillos, S. and Curd, M. (eds.),*Routledge Companion to the Philosophy of Science*, London, Routledge, 115-29.

Horwich, P. 1982, *Probability and Evidence*. Cambridge, Cambridge University Press.

Howson, C. 2003, *Hume's Problem: Induction and the Justification of Belief,* Oxford, Oxford University Press.

Howson, C. y Urbah, P. 2006, 3ª ed., *Scientific Reasoning. The Bayesian Approach,* La Salle, Illinois, Open Court.

Huber, F. 2008, 'Milne's Argument for the Log-Ratio Measure', *Philosophy of Science*, vol. 75, 413-420.

Iranzo, V. y Mtz. de Lejarza, I. 2010, 'Medidas de apoyo evidencial: Un análisis comparativo', *Teorema* Vol. XXIX/3, pp.5-30

Iranzo,V. y Mtz. de Lejarza, I. 2012 (pendiente de publicación) , 'On Ratio MeasuCritical Remarks on Zalabardo's Argument for Likelihood-Ratio Measure', *Journal for General Philosophy of Science*.

Jeffrey, R. 1992, *Probability and The Art of Judgment*, Cambridge, Cambridge University Press.

Jeffrey, R. 2004, *Subjective Probability: The Real Thing*, Cambridge, Cambridge University Press.

Joyce, J. 2004, 'On the Plurality of Measures of Evidential Relevance', *Meetings of the Philosophy of Science Association*; University of Western Ontario

Joyce, J. 2005, 'How probabilities reflect evidence' , *Philosophical Perspectives, 19, Epistemology*, 2005, 153-178

Joyce, J. 2008, 'Bayes' Theorem', *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2008/entries/ bayes-theorem/

Keynes, J. 1921, *A Treatise On Probability*, Londres, Macmillan.

Kuipers, Th. 2000, *From Instrumentalism to Constructive Realism*, Dordrecht, Kluwer.

Milne, P. 1995, 'A Bayesian Defence of Popperian Science?', *Analysis*, vol. 55, 213-215.

Milne, P.1996, 'log *[p (h/eb)/p (h/b)]* Is The One True Measure of Confirmation', *Philosophy of Science*, vol. 63, 21-26.

Rosenkrantz,R.D. 1977, *Inference, Method and Decision*, Dordrecht, Reidel.

Schlesinger, G. N. (1995), 'Measuring Degrees of Confirmation', *Analysis*, vol. 55, 208-212.

Schupbach, J. & Sprenger, J. 2011, 'The logic of explanatory power', *Philosophy of Science*, 78, 105-127

Swets, J.A. 1995, *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers* Lawrence Erlbaum Associates.

Tanner,W.P. y Swets, J.A. 1954, 'A decision-making theory of visual detection' *Psiological Review,*61, pp.401-409

Tentori, K., Crupi, V., Bonini, N., y Osherson, D. 2007, 'Comparison of confirmation measures', *Cognition*, vol. 103, pp. 107-119.

Zalabardo, J. 2009, 'An argument for the likelihood-ratio measure of confirmation' *Analysis*, 69, 630-635.

UN MARCO CONCEPTUAL PARA EL REALISMO ESTRUCTURAL

THOMAS MEIER
Ludwig-Maximilians-Universität
Facultad de Filosofía, Filosofía de la Ciencia y Ciencias de la Religión
Munich Center for Mathematical Philosophy
Munich, Alemania
Thomas.Meier@lrz.uni-muenchen.de

ABSTRACT: The purpose of this work is to outline a new approach to the debate on structural realism. This position has been criticized in the recent literature. The framework of structuralist meta-theory is introduced in order to argue as follows. Once this framework is adapted, it is possible to respond to the critiques that have been made to structural realism and to clarify this position, especially the notion of structural continuity.
KEY WORDS: Structural realism; stucturalist meta-theory; structural continuity; inter-theoretical relations

Introducción

La posición del realismo estructural, expuesta de Worrall, afirma lo siguiente: Cuando nuestras teorías empíricas cambian en el transcurso del tiempo, lo que sí es retenido es su contenido estructural. La estructura matemática de nuestras teorías es lo que se retiene y lo que es continuo, aún en casos de cambio teórico radical. Así, afirma Worrall, hay continuidad estructural entre teorías radicalmente diferentes. Concluyentemente, las partes estructurales de nuestras teorías no se abandonan porque son correctas, mientras otras componentes de las teorías, como los nombres de los referentes, cambian.

En este trabajo, la propuesta es contribuir a una clarificación de la postura original del realismo estructural, proponiendo un marco conceptual adecuado para ello. Este marco conceptual será el de la metateoría estructuralista. La propuesta en este trabajo es mostrar que el enfoque de Worrall no puede abarcar todos los cambios estructurales que ocurren en todas las disciplinas científicas. Se sugiere que no sólo ecuaciones pueden reaparecer o ser retenidas cuando ocurre un cambio teórico radical. Para ello, argumentaré que la concepción de lo que entendemos por una teoría empírica debe ser enfocada desde un punto de vista modelo-teórico. Sólo de esta manera se obtiene la estructura lógica de nuestras teorías empíricas de manera exhaustiva.

1. Realismo estructural – la noción de continuidad estructural

En su trabajo original, Worrall fundamenta su posición en un estudio de caso de la óptica, el cambio de la óptica de Fresnel a la óptica de Maxwell. La continuidad estructural, identificada por Worrall, se manifiesta en que ciertas ecuaciones diferenciales que fueron parte de la teoría de Fresnel, también aparecen como parte de la teoría de Maxwell. En cambio, el término "éter" ya no aparece en la óptica de Maxwell, donde se introduce el campo electromagnético.

> A pesar de que Fresnel estaba bastante mal sobre qué es lo que oscila, estaba correcto, desde el punto de vista posterior, no sólo sobre los fenómenos ópticos, sino también que estos fenómenos dependen de las oscilaciones de algo en un ángulo recto con la luz. Así que, si nos restringimos al nivel de las ecuaciones matemáticas - no notando el nivel fenomenal - hay de hecho continuidad completa entre las teorías de Fresnel y de Maxwell (Worrall, 1989: 118-119).[1]

Lo que Worrall afirma es que las partes estructurales de nuestras teorías empíricas se retienen. Siempre se puede tener conocimiento de las estructuras matemáticas que forman parte de nuestras teorías empíricas. Si ocurre que las mismas ecuaciones son parte de teorías radicalmente diferentes, es razonable asumir que lo que representa el mundo de la mejor manera son estas ecuaciones. Hay dos presunciones metafísicas centrales en el realismo estructural:

(i) Lo que sobrevive cambios teóricos radicales es lo que realmente está aferrado al mundo. Si ciertas estructuras son retenidas durante un cambio teórico, es razonable afirmar que la persistencia de estas estructuras refleja en cada caso específico una parte específica del mundo.

(ii) La persistencia de ciertas estructuras durante cambio teórico hace razonable asumir que al menos la parte estructural de nuestras teorías empíricas representa el mundo correctamente. Por lo tanto, podemos decir que lo que conocemos del mundo son las estructuras mediante las cuales obtenemos una representación correcta de él.

2. Objeciones al realismo estructural

La propuesta de Worrall llevó a una serie de objeciones. Aquellas objeciones han sido presentadas de la manera más representativa por Psillos (1999: 146-161). La siguiente lista expone estas objeciones de Psillos:

(1.) El realismo estructural debe ser distanciado más claramente del antirrealismo. Sólo apelar a la idea de la continuidad a nivel de las ecuaciones matemáticas no basta, ya que también es combinable con el instrumentalismo.

1 Todas las citas en este artículo han sido traducidas del inglés por parte del autor de este artículo.

(2.) No es claro que del hecho que ciertas ecuaciones son retenidas durante cambio teórico se sigue que estas ecuaciones nos dicen algo sobre la estructura del mundo. Especialmente, no queda claro que tales ecuaciones representen relaciones reales entre entidades físicas que serían, de otra forma, incognoscibles.

(3.) Las descripciones sobre el mundo que se obtienen por nuestras teorías empíricas, no se pueden expresar únicamente con ecuaciones. Algunas presunciones teóricas que no forman parte de ecuaciones también se requieren para justificar el éxito de una teoría empírica.

Otra pregunta abierta para el realismo estructural, formulada por Frigg y Votsis (2011), busca aclarar cómo es la relación formal de correspondencia entre dos estructuras. Si una estructura es retenida y reaparece en una teoría nueva, se debe mostrar en qué consiste la correspondencia entre la estructura de la teoría antecesora y la teoría nueva.

3. Proponiendo un marco conceptual para el realismo estructural

La metateoría estructuralista es una concepción modelo-teórica de teorías empíricas. Su exposición canónica se encuentra en Balzer, et.al. (1987). Se caracterizan las teorías en términos de teoría de modelos. En esta concepción, una teoría empírica consiste de sus modelos, son secuencias de la siguiente forma:

$< D_1, ..., D_m, R_1, ..., R_n >$. Las D_1 son los conjuntos básicos y las R_j son relaciones construidas sobre estos conjuntos. Los D_i contienen lo que se toma por la ontología de la teoría, los "objetos" que son asumidos como "reales" por la teoría. Los R_j son funciones. En teorías empíricas que hacen uso de herramientas cuantitativas, son normalmente funciones que proyectan objetos empíricos a los números reales. Una teoría empírica consiste de los siguientes conjuntos de modelos:

(a) La primera parte es el núcleo K, que consiste de la siguiente tupla: $K = < T, I >$. El así llamado elemento teórico T es un conjunto de modelos que consiste de los conjuntos de los modelos potenciales M_p, modelos parciales potenciales M_{pp}, modelos actuales M, condiciones de ligadura globales GC y vínculos interteóricos globales GL. Aparte de T, I es el conjunto de aplicaciones intencionales.

(b) Un elemento teórico T es a su vez una estructura compleja, tiene la forma de la siguiente tupla: $T = < M_p, M_{pp}, M, GC, GL >$

Argumento que el aparato formal de la metateoría estructuralista se mostrará como el adecuado para aproximarse con más claridad al realismo estructural. En lo que sigue se dará

primero una elucidación informal de lo que significa cada parte de un elemento-teórico. Después se mostrará de qué manera se puede fortalecer el realismo estructural con este marco conceptual introducido y se verá un camino para responder a las objeciones, hechas al realismo estructural.

Un conjunto de modelos potenciales (M_p) fija el marco conceptual general, dentro del cual se caracteriza un modelo actual de una teoría. Todas las entidades que se pueden subsumir bajo el mismo marco conceptual de una teoría dada son miembros de los conjuntos de modelos potenciales de esta teoría. Conjuntos de modelos parciales potenciales (M_{pp}) representan el marco conceptual de datos para la corroboración o refutación de una teoría. Los conceptos en M_{pp} se pueden determinar independientemente de T. Esto significa que términos que son teóricos en los conjuntos de modelos potenciales de una teoría se dejan afuera, dado que los M_{pp} son los modelos de datos. Los conjuntos de modelos que no sólo pertenecen al mismo marco conceptual, sino que también satisfacen las leyes de la misma teoría son los conjuntos de modelos actuales M. Dado que aplicaciones locales de una teoría se solapan en el espacio y en el tiempo, se requieren los conjuntos de condiciones de ligadura globales GC para caracterizarlo. Mediante esta herramienta se expresan los requisitos formales que restringen las componentes de un modelo en dependencia de otras componentes de otras modelos. Estas condiciones de ligadura expresan conexiones físicas entre diferentes aplicaciones de una teoría, son sus relaciones intrateóricas. Los conjuntos de vínculos interteóricos globales GL representan las conexiones entre teorías diferentes, dado que ciertas entidades y leyes pueden ser parte de teorías empíricas diferentes.

En la metateoría estructuralista, se asume que después de una reconstrucción lógica (con el aparato formal aquí introducido) de nuestras teorías empíricas, se obtiene como resultado, aparte de la estructura lógica compleja de la teoría, resultados sobre sus relaciones (los vínculos interteóricos) con otras teorías. Dentro de estas relaciones, se pueden identificar estructuras que aparecen en ambas teorías relacionadas. Los respectivos modelos potenciales de las teorías se pueden relacionar mediante tales relaciones. De esta manera, se puede hacer explícita la noción de continuidad estructural dentro del marco conceptual de la metateoría estructuralista. Mediante estudios de caso concretos, reconstruidos con el aparato de la metateoría estructuralista, se pueden modelar de la manera más exacta las relaciones entre teorías empíricas, estas relaciones pueden contar como relaciones de continuidad estructural, en el sentido del realismo estructural.

Si se argumenta que del hecho de que se puedan mostrar formalmente las relaciones interteóricas no se sigue que estas estructuras sean de alguna forma "reales", se debe recordar lo que he formulado como las dos presunciones metafísicas en las que se basa el realismo estructural. Si

aquellas estructuras permanecen una y otra vez más mediante cambios teóricos radicales, esto nos quiere decir que "algo" de nuestra representación estructural refleja el mundo de manera correcta.

En el trabajo de Sneed (1983) se pueden ver semejanzas con el realismo estructural, aunque la metateoría estructuralista nunca ha sido formulada explícitamente como una posición dentro del debate sobre el realismo científico. En este trabajo, Sneed nos dice lo siguiente:

> Es importante entender que, en la vista estructuralista, teorías empíricas hacen aserciones concretas, descriptivas y falsificables sobre su objeto de estudio. El estructuralismo no es solamente una versión nueva de instrumentalismo, escondiéndose detrás de una nube de notación conjuntista. Del otro lado, los estructuralistas ven las estructuras matemáticas asociadas con una teoría como partes mucho más esenciales de la teoría que las aserciones que ella hace. Las aserciones pueden cambiar durante el desarrollo histórico de la teoría, pero el aparato matemático permanece igual. Así, ambos, el estructuralista y el realista afirmarían que la ciencia empírica hace aserciones descriptivas, pero estarían en desacuerdo hasta donde los científicos empíricos, hablando profesionalmente, "piensan lo que dicen" (Sneed, 1983: 351).

En esta cita de Sneed se encuentra claramente una semejanza a las posturas de Worrall. Aunque Sneed no habla de una continuidad estructural y de una retención de ciertas ecuaciones, dice explícitamente que el aparato matemático de una teoría permanece igual. Es ahí donde veo una relación estrecha entre las propuestas de los dos autores. Aunque la metateoría estructuralista ha sido formulada principalmente como un programa para dar cuenta de la estructura lógica de nuestras teorías empíricas, de la dinámica de la ciencia y dado que se piensa como una postura neutral a debates sobre el realismo[2], pienso que la aceptación de su marco conceptual como instrumento de reconstrucción lógica de nuestras teorías empíricas, combinado con las dos presunciones metafísicas expuestas arriba, lleva a una forma de realismo estructural. El realista estructural se puede ver fortalecido de manera significativa, si adapta el marco conceptual de la metateoría estructuralista.

Psillos tiene razón con partes de sus críticas. Con su crítica, Psillos posibilita un cambio de enfoque como lo propongo yo en este trabajo. No sólo las ecuaciones son continuas, sino ciertas partes de nuestras teorías, caracterizadas en términos de la metateoría estructuralista, dependiendo en cada caso concreto de un estudio de caso. Ya que las teorías son identificadas con estructuras, en el sentido formal expuesto arriba, sus conexiones serán conexiones estructurales. Cuando una ecuación se retiene, es simplemente expresado por un vínculo interteórico entre dos modelos potenciales o actuales de dos elementos teóricos. Ahí habrá que ver cada caso concreto, basado en

2 En conversación personal, esto me han dicho tanto Ulises Moulines, como Wolfgang Balzer. Ambos son cofundadores del programa estructuralista.

su reconstrucción lógica concreta. No sólo las ecuaciones, sino la representación completa en términos de la metateoría estructuralista, nos da una representación adecuada de la estructura lógica de la teoría empírica reconstruida. Aquella teoría, a su vez, nos da una representación de ciertos fenómenos de los que trata. No sólo las ecuaciones, sino toda la teoría, entendida como estructura, nos da una representación estructural del mundo.

A su vez, cuestiones de correspondencia entre dos estructuras se resuelven "automáticamente", ya que dentro del marco conceptual de la metateoría estructuralista, cualquier relación entre dos estructuras es expresable formalmente en términos de teoría de conjuntos.

Para ver con más detalle cómo puede ser una continuidad estructural en términos de la metateoría estructuralista, menciono aquí el concepto de *especialización*. Este concepto es parte del marco conceptual de la metateoría estructuralista, definido en Balzer, Moulines, Sneed (1987: 168-172). Una especialización es una relación entre dos elementos teóricos, definida de lo siguiente:

Si $\mathbf{T} = < \mathbf{M_p, M_{pp}, M, GC, GL, I} >$ y
$\mathbf{T'} = < \mathbf{M_p', M_{pp}', M', GC', GL', I'} >$ son elementos teóricos, entonces:
T' es una especialización de T, si:

(1) $M_p' = M_p$
(2) $M_{pp}' = M_{pp}$
(3) $M' \subseteq M$
(4) $GC' \subseteq GC$
(5) $GL' \subseteq GL$
(6) $I' \subseteq I$

La especialización expresa el hecho que nuestras teorías tienen aplicaciones más especificas a un sistema más local. Algunas de las leyes de la teoría en cuestión son aplicadas de manera más especial, o ciertas leyes auxiliares son aplicadas. Esto es expresado en la definición, donde en el punto (3) se expresa que los conjuntos de modelos actuales M' del elemento teórico T' son subconjunto de los conjuntos de modelos actuales M del elemento teórico T. Dado que las leyes de una teoría son partes de los modelos actuales, esto expresa formalmente que las leyes del elemento teórico T' son más especificas que las leyes del elemento teórico T. Mientras tanto, ambos elementos teóricos tratan de lo mismo, dado que sus marcos conceptuales generales, los modelos

potenciales (M_p), son idénticos. La siguiente cita ilustra la necesidad de la introducción del concepto de especialización:

> Cuando consideramos nuestra reconstrucción de la mecánica clásica de partículas, el lector se puede haber preguntado ya donde hemos dejado leyes tan importantes de la mecánica clásica de partículas como la tercera ley de Newton (el principio de actio-reactio), la ley de la gravitación, o la ley de Hook. Nuestra respuesta es: Todos ellos constituyen elementos teóricos diferentes de la mecánica clásica de partículas, pero sí están relacionados. Todo el orden, en cambio, constituye lo que llamamos la "red teórica de la mecánica clásica de partículas". Lo mismo vale para otras teorías avanzadas de la ciencia empírica. … Muchas de estas leyes especiales de la teoría son, más bien, asociados con condiciones de ligadura particulares y, posiblemente, vínculos interteóricos, aparte de los que ya se han explicado, cuando se trata de los elementos teóricos básicos. En otras palabras, la consideración de todos estos requisitos adicionales terminará en la reconstrucción de toda una serie de elementos teóricos diferentes, los cuales, como quiera, tienen la misma estructura básica. Por esta similitud de estructura podemos hablar de una red teórica y no sólo de un conjunto amorfo de elementos teóricos aislados y solitarios (Balzer, Moulines, Sneed, 1987: 168).

El ejemplo de las redes teóricas ilustra de manera ejemplar cómo es que haya continuidad entre dos estructuras. Cuando Worrall se refiere a la continua aparición de ciertas ecuaciones en diferentes teorías, la metateoría estructuralista nos proporciona una noción de continuidad estructural mucho más universal y abstracta. En cada reconstrucción particular de una teoría, la red-teórica tendrá su forma específica, dependiendo de las leyes de las teorías en cuestión. En este sentido, cuando una teoría tiene el suficiente desarrollo y grado de sofisticación, siempre se identificará alguna continuidad estructural en el sentido de especialización dentro de una red teórica. Que aquellas continuidades entre dos estructuras sean informativas, no triviales y que aclaren presunciones del realismo estructural dependerá de la forma concreta de cada red teórica en cada caso concreto.

Conclusión

Pienso haber mostrado que la herramienta adecuada para caracterizar de manera más clara y precisa lo que es el realismo estructural es la metateoría estructuralista. Se ha visto que la noción de continuidad estructural puede ser entendida de una manera más específica. A la vez queda claro que sólo mediante estudios de caso concretos, las posturas del realismo estructural encontrarán su juicio donde se confirmarán o rechazarán. En este sentido, no ha quedado fortalecido el realismo estructural directamente, pero se ha señalado un camino que puede ser recorrido por los realistas estructurales si buscan fortalecer y defender su posición, ya que la propuesta original de Worrall se

ha visto demasiado susceptible a las criticas mencionadas. El realismo estructural como posición metafísica es defendible, de la mejor manera si se adapta el marco conceptual expuesto aquí.

Agradecimientos:
Este trabajo se ha realizado, gracias a la generosa financiación de la Fundación "Alexander-von-Humboldt"

REFERENCIAS

Balzer, W., Moulines, C. U., Sneed, J.M. 1987, *An Architectonic for Science,* Dordrecht, Reidel.
Frigg, R., Votsis, I. 2011, 'Everything you Always Wanted to Know about Structural Realism but Were Afraid to Ask', en *European Journal for Philosophy of Science,* vol. 1(2), 227-276.
Moulines, C. U. 2011, 'Cuatro tipos de desarrollo teórico en las ciencias empíricas', en *Metatheoria. Revista de Filosofía e Historia de la Ciencia*, vol. 1, 2, 11-27.
Psillos S. 1999, *Scientific Realism: How Science Tracks Truth,* London, Routledge.
Sneed, J. 1983, 'Structuralism and Scientific Realism', en *Erkenntnis,* vol. 19, 345-370.
Worrall, J. 1989, 'Structural Realism: The Best of Both Worlds?', en *Dialectica,* vol. 43, 99–124.

INFORMACIÓN BIOMOLECULAR CON CONTENIDO SEMÁNTICO. UNA PERSPECTIVA SISTÉMICA*

ELSA MURO
Universidad de Navarra

MARÍA CEREZO
Universidad de Murcia

ABSTRACT: In this work we present the two interpretations of the explanatory role of the concept of genetic information, namely coding relation and teleosemantics, and the *explananda* to which they contribute. By focusing on an organizational theory of biological functions (Schlosser 1997), we suggest a possible role of the concept of genetic information in the explanation of a third *explanandum*: the continuous self-re-production of an organism by means of which the organism achieves its self-maintanance through time.
KEY WORDS: genetic information, scientific explanation, teleosemantics, biological function

Introducción

El reciente debate en filosofía de la ciencia acerca del rol explicativo que juega el concepto de información en Biología enfrenta a los que consideran que el concepto sólo supondría una contribución teórica entendido como "coding relation" (Godfrey-Smith 2000 y 2008, Jantzen y Danks 2008, Kjosavik 2007), con los que han defendido que el concepto es usado en Biología con una connotación semántica, representacional (Maynard Smith 2000, Shea 2007 y 2012). En este trabajo presentamos las dos líneas generales, que localizan el rol explicativo del concepto en dos tipos de explicación distintas, y acerca de dos *explananda* diferentes. Partiendo de una teoría organizacional de funciones (Schlosser 1997), exploramos un posible rol del concepto de información genética en la explicación de un tercer *explanandum*: la continua auto-re-producción de un organismo con la que logra su auto-mantenimiento a lo largo del tiempo.

En la primera parte, exponemos una síntesis de algunas propuestas contemporáneas del rol explicativo de la noción de información genética, prestando especial atención a la reciente propuesta de Shea (2007, 2012). En la segunda parte, presentamos la concepción de los sistemas vivos que atiende a la dinámica organizacional, esencial para su auto-mantenimiento y auto-re-

* Este trabajo ha sido posible gracias a la financiación del Ministerio de Ciencia e Innovación y del Ministerio de Economía y Competividad del Gobierno de España (REF. FFI2009-13687-C02-01/FISO).

producción, y dentro de este esquema conceptual y científico, introducimos la teoría organizacional de las funciones biológicas, en particular, la propuesta por Schlosser. Aplicamos la propuesta de Schlosser al caso de los genes y apuntamos un concepto de información semántica que tendría su rol explicativo en relación con el *explanandum* de la auto-re-producción.

1. El debate contemporáneo del rol explicativo de la noción de información genética

Anteriores propuestas han identificado dos roles explicativos de dos sentidos distintos de información genética acerca de dos *explananda* diferentes:

A) *Relación codificadora y mecanismo general de síntesis de proteínas*

Godfrey-Smith, Kjosavic y Jantzen y Danks han explorado ciertas características del mecanismo celular de síntesis de proteínas que lo hacen análogo a procesos de transmisión de señales en telecomunicaciones (Godfrey-Smith 2000, Kjosavik 2007, Jantzen & Danks 2008). Han argumentado que la adscripción de propiedades informacionales al ADN contribuye realmente a resolver un problema específico de Biología Celular acerca de la habilidad de la célula para ordenar secuencialmente aminoácidos en la síntesis de proteínas (Godfrey-Smith 2000). En este sentido, el concepto sólo tendría poder explicativo entendido como *relación codificadora* ("coding relation"), pero no conllevaría entender los genes como portadores de contenido semántico. La contribución teórica que el concepto de información genética aporta no entraña directamente ni requiere ninguna hipótesis acerca de la historia evolutiva del organismo (Godfrey-Smith 2000), ni tampoco ninguna adscripción funcional o teleológica que vaya más allá de la generación de las proteínas: se limita a describir las interacciones entre moléculas a través de un tipo especial de causación que es afín a los procesos de codificación de mensajes en la comunicación humana.

B) *Propiedades semánticas y desarrollo de un fenotipo adaptado*

Por su parte, Maynard Smith y Shea, entre otros, han argumentado que el genoma tiene además propiedades *semánticas* que le confieren poder explicativo en explicaciones *funcionales* (teleológicas) de la causación génica y su responsabilidad causal en el desarrollo de fenotipos adaptados (Maynard Smith 2000, Shea 2007 y 2012). Para ellos, las diferentes secuencias de aminoácidos de cada genoma portan un mensaje con contenido intencional, semántico, que ha sido generado en tiempo filogenético, por los mecanismos evolutivos de selección natural, adaptación y transmisión genética, y que el organismo recibe pasivamente y lee a lo largo de su desarrollo.

Por un lado, la comprensión del carácter informacional del DNA como relación codificadora (rol A) atiende a un aspecto organizacional de la causación génica (ordenación de aminoácidos en

las células para conformar proteínas de distintos tipos) pero no presta suficiente atención al hecho de que esa organización molecular a nivel celular tiene una función esencial en la organización y autoconstrucción del sistema total. El ADN tiene un rol en la organización del sistema vivo que va más allá del proteoma, pues al especificar el proteoma de cada célula determina el destino y comportamiento de ésta y su función en el complejo proceso de autoconstrucción del organismo. La expresión diferencial del ADN es una de las claves que explican el desarrollo y funcionamiento de un sistema vivo, que requiere distintos comportamientos de sus células dependiendo del espacio y el tiempo en que éstas se hallen. No todos los genes del ADN idéntico de un organismo se expresan en todas las células ni al mismo tiempo. La mayor parte de los genes se expresan o no dependiendo de la función de la célula en un tejido particular (especificidad espacial) y de los diferentes momentos del ciclo celular y del desarrollo (especificidad temporal).

Por otro lado, la propuesta presentada por Maynard Smith y Shea (rol B) entiende que, además de la relación codificadora, la noción de información genética tiene un rol explicativo en el desarrollo, de nuevo en cada generación, de una complejidad adaptada. En síntesis, la propuesta incluye dos ideas: a) los genes no sólo actúan como codificadores del orden de aminoácidos, sino que son portadores y transmisores de información con contenido semántico en virtud de su función biológica más allá de la ordenación de aminoácidos en la célula (teleosemántica); b) la función biológica de un gen equivale al efecto por el que ha sido seleccionado a lo largo de la evolución (teoría etiológica de funciones).

En este sentido, el contenido representacional de los genes habría sido generado en tiempo filogenético a través de los mecanismos de adaptación, herencia y selección natural. El genoma representaría, así, un contenido imperativo del tipo: hacer un fenotipo adaptado a un ambiente determinado E (Maynard Smith, Shea); y , con él, también un contenido indicativo del tipo: se da tal ambiente determinado E (Shea). En consecuencia, el genoma tendría propiedades semánticas tales como condiciones de satisfacción y corrección respectivamente.

Entre las distintas críticas que las teorías etiológicas han recibido, para nuestros propósitos conviene destacar la que incide en el hecho de que no reflejan el modo en que los biólogos adscriben funciones. En parte de la práctica de la Biología el concepto de función está íntimamente asociado a la contribución que determinados elementos, estados o procesos hacen a alguna capacidad del sistema en que están insertos, y no hace referencia a ninguna hipótesis sobre la evolución ni a aquello que explica su existencia (Christensen and Bickhard 2002, Admunson and Lauder 1994, Walsh 1996, Schlosser 1997, Mossio el al. 2009). Además, la adaptación de un sistema al medio no es un logro que dependa de genes aislados (que raramente se correlacionan con

una diferencia fenotípica) sino más bien de la coordinación entre los distintos genes de cada ADN, los distintos estados del ADN en tiempos y espacios particulares, y otros elementos del sistema que juntos dan lugar a rasgos fenotípicos finales.

En definitiva, a la propuesta semántico-seleccionista se le pueden aplicar las críticas clásicas al programa adaptacionista, es decir, que considera a la selección natural como único agente optimizador del sistema, y que descompone al organismo en unidades y propone una historia adaptativa para cada una de esas unidades tomada aisladamente, cuando un organismo es sujeto de relaciones funcionales externas como un todo en el que no cuenta tanto las partes tomadas por separado, como su integración funcional en el sistema (Gould y Lewontin 1979).

Shea ha formulado un esquema conceptual más sofisticado que el de Maynard Smith. Su estrategia ha sido aplicar una versión de la teleosemántica de Millikan al caso del genoma, concediendo un papel más relevante al sistema consumidor (receptor) a la hora de dar cuenta del carácter semántico de una señal. Shea describe los contenidos representacionales como propiedades relacionales de unos estados físicos denominados "intermedios", por medio de los cuales un mecanismo productor de señales (que puede ser el mundo) y un sistema consumidor (el organismo en desarrollo) interactúan gracias a que el sistema consumidor reacciona sistemáticamente a cada tipo de intermedio diferente, y a que cada una de estas respuestas tiene una "condición de éxito evolutivo", la cual fija el contenido semántico de esos particulares físicos intermedios.

La propuesta de Shea tiene dos consecuencias importantes. Por un lado, rechaza la idea de que la noción de representación genética pueda ser usada para explicar el curso del desarrollo de un organismo. El proceso causal interno que da lugar al desarrollo, según Shea, se explica en términos de las interacciones entre DNA, maquinaria celular y factores ambientales que dan lugar al fenotipo maduro del organismo, sin que pueda hablarse en este nivel de condiciones de satisfacción y corrección que permitan hablar de propiedades representacionales. Por otro lado, y como consecuencia de lo que se acaba de decir, sólo atribuye propiedades representacionales a unos pocos genes, los genes P, es decir, aquellos que se correlacionan con alguna diferencia fenotípica propia de la especie, y deja sin carácter representacional a los demás genes (genes D).

2. El rol del concepto de información genética en la explicación de la auto-re-producción del organismo

En la concepción de los organismos vivos como sistemas auto-mantenidos y auto-organizados que asumimos en este trabajo destacan tres caractarísticas:. Primero, la clausura

organizacional: hay una circularidad causal entre la dinámica microscópica (nivel bajo) y la estructura o patrón macroscópico (nivel alto). Esto implica que la actividad del sistema es necesaria para el mantenimiento del sistema mismo (Mossio et al. 2009). En segundo lugar, la diferenciación organizacional: los diferentes componentes del sistema han de contribuir de modo diferente al auto-mantenimiento del mismo (Mossio et al. 2009). En tercer y último lugar, la auto-re-producción a lo largo del tiempo, lo que (debido a la clausura organizacional, por una parte y a la diferenciación, por otra) conlleva a la vez convergencia y divergencia funcional (multifuncionalidad o funciones cambiantes) (Schlosser 1998).

En este esquema conceptual y científico, la funcionalidad de un rasgo es concebida como algo relacionado no tanto con la historia evolutiva del rasgo, como con su integración en un tipo especial de organización, a cuyo mantenimiento y reconstrucción continua ha de contribuir. Así, un componente funcional de un sistema es, en esta perspectiva, aquel que con su actividad causal condiciona la dinámica del sistema en una dirección determinada, apropiada para que el sistema se mantenga. Por las características especiales de la actividad causal del ADN, y su especial contribución a la capacidad de auto-re-producción del sistema, adoptamos la propuesta de Schlosser, según la cual:

Fc es una función de Xc(t) si y sólo si:

Para un cierto periodo de tiempo $t_0 < t+x+y < t_0 + T$

(1) X(t) es causalmente necesario (directamente) para establecer F(t+x) (bajo ciertas circunstancias c1)

(2) F(t+x) es causalmente necesario (indirectamente) para establecer X (t+x+y) (bajo ciertas circunstancias c2)

(3) Las relaciones causales entre X (t), F(t+x), X (t+x+y) son complejas.

Se trata de una concepción de las funciones biológicas como contribuciones causales a un sistema vivo, que es entendido como un sistema en continua auto-re-producción (1 y 2) y complejo (3). Por un lado, las adscripciones funcionales sólo se pueden hacer a rasgos de un sistema que está *continuamente auto-re-produciéndose* (que pasa a través de secuencias cíclicas de estados y con ello se mantiene estable a largo plazo). Por otro lado, las adscripciones funcionales sólo se pueden hacer a rasgos de un sistema *complejo*. Un sistema complejo es aquí entendido como un sistema que

no pasa de una manera determinista por ciclos simples de estados (como lo hace un planeta), sino que está organizado como una red compleja de distintas transiciones alternativas de estado (todas las cuales convergen siempre en la auto-re-producción del sistema). Dependiendo de las circunstancias del ambiente intra y extraorganísmico, cada estado o rasgo X tiene distintas funciones, es decir, tiene que actuar de diferentes modos, para conseguir la auto-re-producción del sistema. Por ejemplo, un gen X en determinadas células y momentos del ciclo celular tiene que permanecer silenciado, mientras en otras circunstancias es necesario que se exprese, y ambas actuaciones son igualmente necesarias para la auto-re-producción del sistema, y del gen mismo.

El sistema, así, se caracteriza por la divergencia (posibilidad de distintas alternativas de cambio de estado) y la convergencia (contribución a la auto-re-producción del sistema). En este esquema, la funcionalidad de un rasgo depende no sólo de su propia actividad causal sino de que ésta se despliegue en las circunstancias adecuadas, es decir, en un estado particular del sistema. No se puede hablar de funciones "aisladas" (como hacen las teorías etiológicas) sino de funcionalidad condicionada a las circunstancias, una funcionalidad que sólo es tal en la integración con la organización cambiante del sistema. Se trata de una funcionalidad "condicional": la necesidad de auto-re-producción del sistema complejo bajo ciertas circunstancias. La funcionalidad, así, no es una propiedad "intrínseca" de un rasgo o proceso, sino una propiedad "relacional", en el sentido de que refiere a la relación entre el rasgo y los estados circunstanciales del ambiente (ambiente tanto externo como interno del sistema).

La divergencia es constitutiva tanto de la clausula (1) (un mismo rasgo es capaz de diferentes funciones dependiendo de c1) como de la clausula (2) (F puede ser necesario para la auto-re-producción del sistema y de X sólo bajo determinadas circunstancias c2). Nos interesa ahora centrar la atención en la divergencia en (2): un gen X sólo contribuye con su expresión (con la síntesis de la proteína que codifica) a la auto-re-producción del sistema si está en determinadas circunstancias tanto espaciales (en qué tipo de célula se encuentra) como temporales (en qué etapa del ciclo celular, y del desarrollo del sistema, se halla).

En la auto-re-producción de un sistema vivo juega un papel esencial la diferenciación celular, que está determinada por la síntesis diferencial de proteínas en cada célula (que, determinada por la expresión diferencial del ADN, determina la función de esa célula en el organismo) y por la reproducción celular (iniciada con la replicación del ADN, que es dependiente también de las proteínas sintetizadas). La síntesis de proteínas en cada célula se lleva a cabo a partir de i) un ADN que es estructuralmente idéntico en todas las células del cuerpo y ii) unos mecanismos biomoleculares de selección de las secuencias de ese ADN que han de ser transcritas.

Dependiendo de las circunstancias del ambiente inter e intracelular, cada ADN ha de expresar unos u otros genes para mantener al sistema.

Si aplicamos la definición de Schlosser al caso del genoma, podemos decir:

La síntesis de un tipo particular de proteína (P) es una función de una secuencia de ADN (G) en determinadas circunstancias y en un tiempo t si y sólo si

Para un cierto periodo de tiempo $t_0 < t+x+y < t_0+T$

(1*) G es en t causalmente necesaria (de un modo directo) para sintetizar P en (t+x) (bajo ciertas circunstancias c1).

(2*) La síntesis de la proteína P es causalmente necesaria para que (de un modo indirecto, y en unas determinadas circunstancias c2) el sistema continúe auto-re-produciéndose, y con él, re-produciendo también a G en (t+x+y) (por *replicación* del ADN).

(3*) Las relaciones causales entre G en t, la síntesis de P en (t+x), G en (t+x+y) son complejas.

A partir de lo expuesto hasta ahora, entendemos que hay un tercer *explanandum* que también podría ser explicado con un concepto de semántica génica, y que no queda explicado por los genes semánticos de Shea ni de Maynard Smith:

Explanandum C: auto-mantenimiento del sistema vivo a lo largo de las sucesivas transiciones de estado a través de las cuales se auto-re-produce continuamente desde que nace hasta que muere.

Nuestra propuesta trata de dar explicación a la auto-re-producción del sistema por medio de la atribución de propiedades representacionales a los genes D. En este caso, los genes D tienen como contenido indicativo el darse de un particular estado o condición del ambiente intra e intercelular (las circunstancias c2 en las que la expresión de ese gen es funcional, contribuye a la auto-re-producción del sistema); y como contenido imperativo el que dicta hacer la proteína que el organismo necesita para poder seguir auto-re-produciéndose.

Las anteriores propuestas semánticas de la información genética, y en especial la de Shea, han dado con un sentido en el que el genoma es portador de representaciones en virtud de las propiedades relacionales de estados del ADN (los genes) con el ambiente externo y la evolución de la especie. En nuestra propuesta, el genoma es asimismo portador de representaciones en virtud de las propiedades relacionales de estados del ADN (los genes) con las circunstancias *internas* del sistema complejo auto-re-productor.

En general, tanto para la perspectiva etiológica como para la organizacional, la funcionalidad de un gen estriba en que dirige la actividad del sistema en una dirección que le beneficia. De igual modo que Shea reconoce un contenido representacional a los genes P en la medida en que sus efectos son funcionales *si* se da una condición del ambiente externo al sistema, podemos decir, por el carácter esencialmente relacional (condicional) de las funciones desde la perspectiva organizacional, que los genes D son funcionales *si* se da una condición del ambiente interno al sistema (c2). Ese gen D (o el ARN resultante del mismo) sería, así, portador de un contenido semántico indicativo del tipo: se da c2, pues condicionaría el comportamiento celular en una dirección que es beneficiosa si se da c2.

REFERENCIAS

Godfrey-Smith, P. 2000, `On the theoretical role of `genetic coding´´, *Philosophy of Science* 67, 26-44.

Gould, S. J. y Lewontin, R.C. 1979, `The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme´, in *Proceedings of the Royal Society of London, Series B, Biological Sciences* 205, 581-598.

Jantzen, B. y Danks, D. 2008, `Biological codes and topological causation´, *Philosophy of Science* 75, 259–277.

Kjosavik, F. 2007, `From symbolism to information? – Decoding the gene code´, *Biology and Philosophy* 22, 333–349.

Maynard Smith, J. 2000 `The concept of information in Biology´, *Philosophy of Science* 67, 177-194.

Mossio, M., Saborido, C. y Moreno, A. 2009, `An organizational account of biological functions´, *British Journal for the Philosophy of Science* 60, 813-841.

Schlosser, G. 1998, `Self-re-Production and Functionality: A systems-theoretical approach to teleological explanation', *Synthese* 116, 303–354.

Shea, N., 2007, `Representation in the genome and in other inheritance systems´, *Biology and Philosophy* 22, 313–331.

Shea, N. 2012, `Inherited representations are read in development´, *British Journal for the Philosophy of Science* 63, 1–31.

CONCEPTIONS OF MECHANISMS AND INSENSITIVITY OF CAUSATION

VIOREL PÂSLARU
University of Dayton
vpaslaru1@udayton.edu

ABSTRACT: Conceptions of mechanisms due to Glennan (1996; 2002), Machamer, Darden, and Craver (2000), Bechtel and Abrahamsen (2005) have developed in opposition to the nomological approach to explanation. It is less emphasized, however, that these conceptions have also developed as alternatives to the causal perspective on explanation. In this paper, I argue that despite their distancing from the topic of causation, the mechanistic conceptions need to incorporate in their definitions of mechanisms the notion of insensitivity of causal relations that was examined by Woodward (2006).
KEYWORDS: mechanism; causation; insensitivity.

1. Insensitivity of Causation

Woodward (2006) claims that a causal relationship '*C* causes *E*' is sensitive if it holds in the actual background circumstances *Bi*, but fails to hold in the circumstances *Bii-n* that depart in various ways from the actual circumstances. A causal relationship is less sensitive, *i.e.*, more insensitive or stable, if it holds over a broader range of circumstances that depart in various ways from the actual ones. Suppose Suzy throws a rock at a glass bottle and it shatters. The causal relationship between Suzy throwing the rock and the bottle shattering is insensitive to various changes in the background conditions in the sense that it will continue to hold even if the throwing and the shattering occur in Boston while someone sneezes in Chicago, regardless of the color of Suzy's blouse and despite variations in environmental conditions (temperature, wind speed, etc.). The bottle would have shattered had Suzy thrown the rock earlier or later, or from a different angle. If departures from actuality alter the time and place of the causal relationship, this may affect the identity of causal relata such that instead of a *C*-event and an *E*-event there will be a *C*-like and an *E*- like event (Woodward 2006). This would probably translate into a case involving Jenny in Chicago, instead of Suzy in Boston, throwing a metallic object, instead of a rock, at the bottle that subsequently shatters.

Woodward's proposal can be examined more specifically to suggest three types of insensitivity to changes in the structure of a causal relation and two types of background conditions. *Componential insensitivity* of a causal relationship obtains when the relationship is not affected by

changes in its relata. Suzy might throw a rock, or a brick, or a metallic object at the bottle and it will break, preserving thereby the causal relationship. *Organizational insensitivity* characterizes the causal relationship that is not altered by changes in the arrangement of its relata. Suzy might throw the rock from various distances or under various angles, but the relationship will be preserved, since the bottle will break. *Interactional insensitivity* characterizes the case when changes in how the relata interact do not impact the causal relationship. Suzy could throw the rock more or less forcefully, but this will not affect the relationship as long as the bottle breaks. However, if Suzy throws the rock with a too small force and the rock does not reach the bottle, the relationship breaks. It is sensitive to a throw executed with a small force.

The color of Suzy's blouse and the Dow index are background conditions that *do not and cannot* affect the causal relationship rock throwing – bottle breaking. Because of that, exploring them will not provide important information about the relationship. By contrast, wind is an example of background conditions that *do not*, but *can* affect the causal relationship. If wind blows with, say, a 65 km/h speed, it can blow away the bottle, influence the trajectory of the rock, but the rock will nevertheless strike the bottle. This illustrates background conditions that *affect* a causal relationship, yet it still occurs, or is insensitive to influences from such conditions. Due to their potential to influence causal relations, background conditions of the latter two types are relevant for investigating causal relations.

2. Conceptions of Mechanisms and Insensitivity of Mechanisms

In light of the foregoing consideration of insensitivity, I turn now to examining the notions of mechanisms and the examples of mechanisms that Machamer et al (2000), Glennan (2002), Bechtel and Abrahamsen (2005) use to explain their conceptions of mechanisms and mechanistic explanation.

Machamer et al claim that "[m]echanisms are regular in that they work always or for the most part in the same way under the same conditions" (2000, 3). Yet they do not clarify in what respects the conditions are the same. Their claim could be interpreted as indicating insensitivity of mechanisms under background conditions that (a) do not change, or (b) change without affecting the functioning of the mechanism. Version (a) does not seem to do justice to some mechanisms and is, hence, problematic. Version (b) might mean that conditions could be the same in the sense that conditions *Bii* of one neuron differ only by a small degree from conditions *Bi* of another neuron, or even if they differ by a greater degree, they do not contain factors that might affect the mechanism functioning. Conditions *Bi* and *Bii* could be the same in the sense that they both contain the same

ions of Ca, Na, and Mg, but in different quantities. Moreover, the claim does not suggest whether the conditions affect in any way the functioning of the mechanism. If my reading of "the same" is correct, Machamer et al leave out the situation when the conditions are *different*. For example, *Bii* could contain factors, say, molecules of Methylenedioxymethamphetamine, that *Bi* does not, and these factors influence the production of phenomenon, but do not block it. Such cases show that their description of background conditions is incomplete. That "[t]he organization of ... entities and activities determines the ways in which they produce the phenomenon" (Machamer et al 2000, 3) suggests that the phenomenon production is sensitive to changes in the organization of component activities and entities, but they do not consider the case of organizational insensitivity, *i.e.*, when a mechanism produces a phenomenon despite changes in its organization. It seems, however, that they admit componential insensitivity. Along this line could be read the claim "Activities usually require that entities have specific types of properties" (Machamer et al 2000, 3), which leaves open the possibility that any type of entity that has the relevant properties can execute the activity. The examination by Machamer et al of mechanisms does not seem to engage interactional insensitivity.

The mechanisms of chemical transmission at synapses and depolarization illustrate the view of Machamer et al on mechanisms. Examination of these mechanisms shows that they exhibit multiple kinds of insensitivity and the characterization of mechanisms by Machamer et al should explicitly consider them. The following analysis of these mechanisms is based on Levitan and Kaczmarek (1997).

Calcium channels illustrate insensitivity to the influence of background conditions that *affect* casual relations, as well as componential insensitivity. A subunit, α_1, and one, or several, smaller subunits β, γ, α_2, and δ constitute calcium channels. α_1 is sufficient to make a functional voltage-dependent calcium channel. The smaller subunits are not necessary for channel activity, yet they interact with the α_1-subunit and modulate the kinetic properties of the channel. β subunit, when expressed with α_1-subunit, increases the rate of inactivation of the channel. Although the channel inactivates at a faster rate, the β subunit does not block its functioning. The functioning of the channel is thus insensitive to the presence of this subunit. The latter modulates the channel functioning, but does not break it. Another case of insensitivity refers to the interaction between neurotransmitters and receptors. Virtually every neurotransmitter interacts with more than a single class of receptors (Levitan and Kaczmarek 1997, 236). Acetylcholine interacts with nicotinic as well as muscarinic receptors. The insensitivity of the relationship between neurotransmitter and receptors is limited; it breaks if the receptor is a different one, *e.g.*, NMDA. Yet it is insensitive because receptors do not interact with just one type of neurotransmitters. Sensitivity can be found in

the relationship between transporter proteins and a particular neurotransmitter. The mechanisms of neurotransmitter uptake involve distinct transporter proteins that are specific because they transport a particular neurotransmitter (Levitan and Kaczmarek 1997, 220). The relationship between transporter proteins as part of the mechanism of neurotransmitter uptake and neurotransmitters is highly sensitive to the identity of the component neurotransmitter that it is supposed to transport. Change the neurotransmitter, and the transporter protein does not take the neurotransmitter and the uptake mechanism does not function.

Modulation of patterns of electric activity of cells due to the presence of certain chemicals in the cell environment further illustrates insensitivity to changes in background conditions that affect a causal relationship. Action potentials result from the activity of a population of ion channels. Calcium channels contribute to the depolarizing phase of the action potential, while potassium channels underlie the repolarization phase and contribute to determining the action potential shape. The calcium and potassium channels, the ions, the cell membrane are the key components of the mechanism responsible for action potential. The environment of the neuron comprises the background conditions. A variety of neurotransmitters might be present in it. The neurotransmitters might decrease the calcium or sodium currents, narrowing as a result the action potential, as it happens for instance in the neurons of the chick dorsal root ganglion. Alternatively, introduction of noradrenaline, or of other similar agonists, in the environment of cardiac muscle cells increases the activity of calcium channels, which in turn prolongs significantly the cardiac action potential. The neurotransmitters and the noradrenaline modulate the action potentials. The latter would have occurred even if those substances were not present in the cell environment. Given that neurotransmitters and noradrenaline do not break the relationship between ion channels and action potentials, but modify it, the relationship is insensitive. The neurotransmitters and the noradrenaline make up the background conditions that can change a causal relationship, and the relationship holds despite changes in these conditions.

The foregoing neurobiological example does not illustrate organizational insensitivity, yet Machamer et al need to consider it for their account of mechanism to be general.

Bechtel and Abrahamsen have recently offered this characterization of mechanisms:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism, **manifested in patterns of change over time in properties of its parts and operations,** is responsible for one or more phenomena (Bechtel and Abrahamsen 2010, 323).

The phrase in boldface is an augmentation of their earlier (Bechtel and Abrahamsen 2005, Bechtel 2006) characterization of mechanisms. The purpose of augmentation is to extend the mechanistic explanation to accounts of dynamics of mechanisms by means of computational modeling. Neither the older nor the recent characterizations consider the behavior of mechanisms when there is change in their background conditions or in their structure. However, Bechtel and Abrahamsen's discussion of homeostatic or autopoietic organization of organisms can be interpreted as insensitivity to background conditions that can or do affect the mechanism. Homeostatic or autopoietic organization occurs in living systems that "resist those environmental forces that threaten them," as well as in systems that simply maintain themselves even if not threatened by environmental forces (Bechtel 2006, 45). These systems are insensitive to influences from background conditions that can affect them. Bechtel and Abrahamsen do not examine mechanisms that do not have the specific homeostatic or autopoietic organization and yet resist "environmental forces that threaten them," nor do they examine the case of insensitivity to structural changes. The heart example that they use to illustrate their conception of mechanisms presents however sufficient instances of insensitivity to changes in its structure and in its background conditions.

Hearts have features that escape Bechtel and Abrahamsen's characterization of mechanisms. The heart mechanism undergoes changes over time and yet its operations are stable and they produce the phenomenon of blood circulation. The fetal heart shows a rate of 120 to 160 beats per minute, that of an infant presents a rate of 120 beats per minute. The rate slows to 90 rates per minute when the child is seven years old, and stabilizes at 70 rates per minute at age 18. Depending on which activities one engages in, the heart rate would increase or decrease. If rate is an aspect of organization, we have here an instance when the mechanism is insensitive to changes in its organization. The mechanism does not break although the organization changes. Moreover, rate changes mean differences in how components interact, suggesting interactional insensitivity. The heart mechanism is also insensitive to a number of significant changes in its components. Until the heart becomes of an adult, it is of a fetus, infant, toddler, child, and teenager. The components of the heart change; they grow. There are also macromolecular changes in the heart tissue from birth to adulthood. For example, there are changes in the expressed proteins, which affect muscle contractility (Martinsen and Lohr 2005). Thus, the heart is developmentally insensitive to modifications in its structure that occur over long and short periods of time.

A more radical change in the components of hearts presents replacement of defective valves with mechanical ones (from titanium or ceramic), or biological ones (from human, porcine or bovine tissue). Similarly, blocked arteries in some hearts are replaced with coronary bypass grafts.

The functioning of the heart mechanism is insensitive to such changes in its components and organization, and it produces the blood circulation phenomenon.

Transplanted hearts show insensitivity to changes in background conditions. Some of these conditions in the recipient patient, like the thorax, do not affect the functioning of the heart and it is insensitive to changes in thoraxes. Others, for example the immune system of the receiving organism, can and do affect it by rejecting the donated organ. The functioning of the heart mechanism can be, at least for some time, insensitive to changes in immune systems from donor to recipient and to adverse reaction of the latter.

Fetal hearts present a different departure from the adult heart prototype. The interatrial septum of a fetal heart contains the *foramen ovale* valve that allows the blood to be shunned from the right (pulmonary) side to the left (systemic) side. The ligament of the inferior vena cava aids this flow and the duct of the artery connects the left pulmonary artery and the aortic artery. These specific features of the components of fetal hearts are required because fetal blood flows through umbilical atria and arteries, and oxygenation occurs in placenta, as opposed to in the lungs. At birth, the *foramen ovale* closes and the heart chambers are septated, the duct ceases its function, and blood circulates as in adults (Weinhaus and Roberts 2005). In Woodward's terminology, fetal blood circulation is *like* adult blood circulation. Correspondingly, the structure of the fetal heart is *like* that of an adult. Except for the *foramen ovale*, the ligament, the duct and the blood flow, the heart mechanism is insensitive to changes from fetal phase to adult phase. Bechtel and Abrahamsen (2005) somewhat capture this particular aspect of insensitivity when they claim that a mechanism presents *variations* relative to a prototype, their focus being on componential insensitivity, but not on the other types of insensitivity. What does not vary between the prototype and the variant is an insensitive structure.

Glennan's characterization of mechanisms does not show an explicit acceptance of any version of insensitivity:

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interaction between parts can be characterized by direct, invariant, change-relating generalizations (2002, S344).

An important feature of mechanisms is that the arrangement of their parts is stable. Glennan explains first stability of parts: "in the absence of interventions, their properties must remain relatively stable" (Glennan 2002, S345). Stability in this sense is rather similar to the insensitivity of Suzy's throw and the shattering of the bottle given the color of her blouse. The relationship between the throw and the shattering is stable because the blouse does not affect it. By contrast, the

properties of parts could remain stable *despite* interventions, but Glennan does not consider such a case. Watches, cells, organisms and social groups are mechanisms "consisting of stable arrangements of parts" (Glennan 2002, S345). Given the foregoing clarification of "stable," mechanisms turn out to be stable because nothing intervenes on them. However, the mechanisms that Glennan gives as examples suggest a different understanding of stability. A social group is subject to social influences, a watch is subject to various mechanical stresses, and the immune system of the recipient patient affects the heart, and yet all three maintain their stability, which would amount to insensitivity to background changes that affect the mechanism. Since Glennan uses the heart and cell examples as illustrations of his conceptions, the findings concerning these mechanisms that I outlined above hold here as well. For the sake of brevity, I will not repeat them here, but I will remind that they show that causal relations and mechanisms are insensitive to changes in their components, interactions between them and their organization.

Glennan also states that the behaviors of these mechanisms "can manifest themselves at more than one time and place"(2002, S345). This would be a case of insensitivity to changes in the background conditions time and space. However, if the two conditions are not accompanied by other conditions that can affect the functioning of the mechanism, then the capacity of a mechanism to manifest its behavior at various times and places is a rather weak form of insensitivity.

Conclusion

The foregoing examination has shown that insensitivity is implicit to various degrees in the conceptions of Machamer et al, Glennan, Bechtel and Abrahamsen. Their examples of mechanisms are insensitive in many important ways and to different degrees. To account for these cases, their conceptions of mechanisms need to incorporate the notion of insensitivity to structural changes and to background conditions.

REFERENCES

Bechtel, W., and Abrahamsen, A. 2010, 'Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science', *Studies in History and Philosophy of Science Part A* 41, 321-333.

Bechtel, W., and Abrahamsen, A. 2005, 'Explanation: a mechanist alternative', *Studies in History and Philosophy of Biological and Biomedical Sciences* 36, 421-441.

Glennan, S. 1996, 'Mechanisms and the nature of causation', *Erkenntnis* 44, 49-71.

Glennan, S. 2002, 'Rethinking mechanistic explanation', *Philosophy of Science* 69, S342-S353.

Levitan, I.B., and Kaczmarek L.K. 1997, *The neuron: cell and molecular biology*, New York, Oxford University Press.
Machamer, P, Darden L., and Craver C. F. 2000, 'Thinking about mechanisms', *Philosophy of Science* 67, 1-25.
Martinsen, B., and Lohr J. L. 2005, 'Cardiac development', in Iaizzo P.A. (ed.), *Handbook of cardiac anatomy, physiology, and devices*, Totowa, Humana Press Inc., 15-23.
Weinhaus, A. J., and Roberts K.P. 2005, 'Anatomy of Human Heart', in Iaizzo P.A. (ed.), *Handbook of cardiac anatomy, physiology, and devices*, Totowa, Humana Press Inc., 51-79.
Woodward, J. 2006, 'Sensitive and Insensitive Causation', *Philosophical Review* 115, 1-50.

MEDICIÓN EN PSICOLOGÍA. PONDERACIÓN DE LAS CONCEPCIONES CLÁSICAS

DIANA LUZ RABINOVICH
Universidad Nacional de Córdoba - CONICET

ABSTRACT: I describe some of the biases of classical accounts of measurement in psychology, that help to understand the widespread treatment of the problems in this context. Towards this end, I present the crucial distinction between representationalism and operationalism, and some subsidiary distinctions like the one between metric and non-metric or quantitative and qualitative, and between measurement theory and metrization theory. I then show some alternative ways to work on the problem of measurement in psychology, beyond classical accounts. Finally I develop some related key issues and the constraints that the previously drawn distinctions provide for their treatment. I outline some relevant questions for further research, such as the issue of accuracy in the study of qualitative aspects of psychological phenomenon and the relevance of features inherent to mathematical modeling and their role in measurement.
KEYWORDS: Measurement in Psychology, Representationalism, Operationalism, Measurement Instruments, qualitative phenomenon

En el presente trabajo describo algunos de los sesgos de las concepciones clásicas sobre la medición en psicología que facilitan la comprensión del tratamiento generalizado de los problemas en ese contexto. Para ello presento la distinción crucial entre representacionalismo y operacionalismo, y algunas subsidiarias como la entre métrico y no métrico o cuantitativo y cualitativo, teoría de la medición y de la metrización. Presento luego vías alternativas para trabajar sobre la cuestión de la medición en psicología, más allá de las concepciones clásicas. Finalmente muestro algunas problemáticas claves y los constreñimientos que las distinciones antes trazadas constituyen para su tratamiento. Aparecen algunos aspectos de interés para su posterior investigación. Entre ellos el relacionado con la precisión en el estudio de los aspectos cualitativos de los fenómenos psicológicos y el de la importancia de los aspectos propios del modelado matemático y su rol respecto de la medición.

1. Una historia en blanco y negro: representacionalismo y operacionalismo

Tal como lo reconstruye Joel Michell (1999) la historia del concepto de medición en psicología ha seguido un derrotero hacia la evasión de lo que él llama la *objeción de cantidad*. Ésta refiere al hecho de sostener que los atributos psicológicos no sean cuantitativos. Michell propone

que se ha dado este fenómeno por el corrimiento desde una concepción representacionalista a una operacionalista. La concepción representacionalista de la medición es descripta por las posiciones de Russell, Campbell y Nagel. Para el representacionalismo el corazón de la medición ya no descansa en las proporciones de la naturaleza (como lo hiciera para el realismo de Euclides) sino en las representaciones numéricas que puedan ofrecerse para ciertas relaciones empíricas, mediante una correspondencia o mapeo. Esta teoría fue desarrollada maduramente por Suppes en conjunto con Krantz, Tversky y Luce, y se formó a partir de su experiencia práctica con los experimentos en psicología como un enfoque altamente formalizado entre los años 1970s y 1990s. En contraposición, para el operacionalismo definido por la concepción de Stevens (1946) la manera de evitar la *objeción de cantidad* estuvo basada en neutralizar el énfasis representacionalista en la búsqueda de estructuras cuantitativas en los fenómenos. De la mano de esta estrategia conceptual Stevens abrió la puerta para tomar como mediciones de magnitud numérica cualesquiera respuestas que una operación pudiera definir como tales, sin demandar ya al fenómeno ninguna estructura o naturaleza específica. De acuerdo con esta reconstrucción, las teorías sobre la medición pendularían entre la atribución o no de estructuras subyacentes a los fenómenos.

2. El sesgo respecto de las distinciones métrico-no métrico o cuantitativo-cualitativo

Paralelamente al establecimiento de la distinción más arriba introducida se ha trazado una entre conceptos métricos y no métricos (clasificatorios y comparativos). Esta distinción fue propuesta por Carnap (1950) y Hempel (1952). Mientras que los conceptos clasificatorios tan sólo distinguen clases, los conceptos comparativos establecen relaciones de orden. Si bien se pueden asignar números tanto a clasificatorios como a comparativos, lo que los distingue de los conceptos métricos, como notan Díez y Moulines (1997), es que éstos permiten operar aritméticamente sobre los números asignados. Ahora bien, esta asignación de números que ofrece la capacidad de operar sobre ellos con sentido implica una exclusión: el problema que circunscribe es el de la asignación de números. Sin embargo, existen conceptualizaciones operables, es decir matemáticas, que no involucran numeración aritmética. Es decir, entre los conceptos no-métricos, especialmente en los comparativos, uno podría incluir algunos tipos de conceptualizaciones que pueden ser algebraicas, topológicas, geométricas, etc. Estas conceptualizaciones ofrecen un tratamiento de sumo interés que quedaría opacado ante una distinción tajante. Una consecuencia de esto ha sido descripta por Michell en 1999; que la psicología tratando de volverse cuantitativa emuló las ciencias cuantitativas en el sentido superficial del desarrollo de instrumentos que producían outputs numéricos.

El tratamiento que hace Michell sufre del perjuicio de las oposiciones exclusivistas que he descripto. Si bien este autor da cuenta de la importancia de la ampliación del objeto de las matemáticas hacia la inclusión de estructuras no cuantitativas, propone como alternativa al operacionalismo una vuelta al representacionalismo. Esto es, a la búsqueda de cierto tipo de estructura interna y su correlación con una escala de medir adecuada. En este contexto ha sido problemática la integración de aspectos cualitativos. Un caso que muestra un contrapunto es el de las investigaciones basadas en la medición longitudinal (Ruspini 2002), que utilizan tanto elementos de tradición cuantitativa como cualitativa. Estos proyectos muestran que los cuantitativos y cualitativos no son modelos mutuamente exclusivos, pero además que se ve postergada la individuación de una estructura definitoria para el fenómeno.

3. El sesgo respecto de la distinción entre teoría de la medición y de la metrización

José Díez (1997) distingue la teoría de la metrización como la de las condiciones de posibilidad para la medición, de la teoría de la medición que trata la asignación de números a objetos para representar sus propiedades de magnitud capaces de 'más o menos'. La teoría de la metrización es una investigación teórica previa que trata sobre la posibilidad de leyes cualitativas que hacen que un sistema empírico sea medible. Esta investigación refiere a lo que ya Helmholtz sostuvo respecto de que 'deberíamos investigar en cuáles circunstancias podemos expresar magnitudes a través de (...) números' (Helmholz 1921, 89), la respuesta que dio fue que '[l]as propiedades medibles de los cuerpos deben ser similares en alguna manera especial a las propiedades de los números' (Helmholtz 1921, 112). La metrización investiga acerca de si los sistemas satisfacen ciertas condiciones o leyes de la naturaleza que dan sentido a la medición. Díez relata por ejemplo que el artículo clásico de Stevens fue rechazado en base a la inexistencia de una operación aditiva de concatenación para las sensaciones, y a que no era claro si se trataba de una medición derivada de ciertas magnitudes psicológicas o del establecimiento de leyes psicofísicas entre magnitudes medidas independientemente. En todo caso, lo que se requería es que se estableciera una estructura análoga a alguna escala para las sensaciones psicofísicas. Stevens estudió de manera formal las diferentes relaciones que hay entre diferentes escalas para representar una magnitud, pero no dijo nada acerca de por qué estas diferentes representaciones son representaciones de la misma magnitud. En una continuación del artículo Díez (1997) relata como estas dos inquietudes, la de fundamentar la medición y lograr esta sea adecuada, quedan integradas en la teoría de la metrización madura de Suppes. Nuevamente aquí podemos observar el sesgo respecto de que no hay una preocupación auténtica por ir más allá de la identificación de estructuras

definibles formalmente para una cierta medición escalar. A la vez, un lugar meramente accesorio y de carácter técnico (no teórico) queda para la medición propiamente dicha.

4. Más allá del clasicismo. Acerca de los instrumentos de medición

Morgan (1999) aclara el panorama en general, planteando dos alternativas al tratamiento clásico de la medición. Para ella hay tres tipos de literaturas que nos ayudan a pensar analítica y seriamente sobre la historia de la medición (sus estudios se desarrollan para el caso de la economía pero creo que es pertinente para la psicología también, pues se centra en la problemática de medir fenómenos que no son fáciles de de medir). Estas literaturas provienen de la filosofía de la ciencia, de la metrología y de la historia y los estudios sociales de la ciencia. La posición dominante en filosofía de la ciencia es la teoría representacional de la medición que ya he descripto más arriba. Morgan agrega un tinte de interés citando a Finkelstein (1982) quien ofrece una versión más pragmática para investigar e interpretar los requerimientos formales que aseguran la medición válida en la teoría representacional, poniendo énfasis en el rol de la objetividad y en que la medición es un proceso empírico que es el resultado de la observación y no, por ejemplo, de un experimento mental. Esta versión práctica de la teoría representacional nos dejaría en el límite con el segundo enfoque –el de la metrología– desarrollado para la economía por Boumans (2001, 2004, 2005). Él sostiene la importancia de tomar más seriamente la idea de que hay *instrumentos de medición*. Éstos instrumentos, en economía consistirían en fórmulas matemáticas, modelos, e incluso partes de modelos que se usan como dispositivos o instrumentos que permiten hacer mediciones (por ejemplo asignar números) a entidades difíciles de medir. Estos instrumentos pueden ser vistos como teniendo un comportamiento análogo a los de la medición tradicional, pues son independientes de los cambios dados en el medioambiente más amplio pero al mismo tiempo capturan con precisión las variaciones de la entidad que se quiere medir. Sirven para extraer la señal del ruido, y filtrar y calibrar esta señal en números.

En paralelo a estos enfoques Morgan describe como Porter ha desarrollado en la historia de las ciencias sociales una tendencia de análisis focalizado en las maneras en las que los números en ciencias sociales son aceptados como mediciones legítimas y convencionales en sus campos específicos, como *reglas cuantitativas estandarizadas*. Las tres nociones relevadas por Morgan tienen en común la idea de que necesitamos un instrumento de medición, aunque la naturaleza de tales instrumentos (una estructura empírica relacional, una fórmula modelo, o una *regla cuantitativa estandarizada*), y los criterios para su adecuación, hayan sido dispuestos de manera muy diferente.

La idea de Morgan, respecto del campo de la economía, es que puestos en juego estos instrumentos muestran diferentes utilidades y ajustes y dan una mejor idea sobre en qué consiste la medición de hecho. Por caso, los instrumentos representacionales tienen dificultades para ser mapeados con números del mundo económico pues la estructura empírica exige relajar los axiomas de la numérica (Boumans 2001), o bien existen ejemplos donde instrumentos de medición estadísticos confiables han sido construidos en base a teorías que luego se han tornado falsas.

Boumans en 2005 retoma la idea de *instrumentos de medición* (para la economía) y muestra algunos de sus aspectos más ricos para la investigación. Por ejemplo, respecto de la variación y la invariancia en la medición. El instrumento debería ser confiable en diferentes lugares y períodos de tiempo para proveer mediciones comparables (estandarizadas), pero al mismo tiempo deberían capturar las variaciones de manera precisa a través de diferentes lugares y tiempos. Esto implicaría que debe ser posible distinguir cuando la constancia en el tiempo de una medición sugeriría una constante natural o bien si la fórmula no está capturando cierta variación relevante. La ecuación podría estar funcionando meramente como un dispositivo de chequeo o control y no como instrumento de medición. Como lo problematiza Toomela (2008) la entidad medida en ese caso no es en realidad una figura significativa sino sólo una correlación estadística numérica que no muestra los constreñimientos de interés en la naturaleza. Sólo mide una razón entre dos fenómenos medidos de antemano, sin conexiones causales autónomas. Pero, también podría ser el caso, como menciona Morgan, que se está efectuando la medición de un concepto independiente y se exhibe su tendencia, lo que aún generando desconfianza puede ser útil para la predicción. Una alternativa interesante es que tal concepto mantuviese una relación potencialmente confiable y potencialmente analizable con otra entidad medible. Un mundo de aristas se abren al análisis.

5. Más allá del sesgo. Sobre causas y estadísticas

Que la medición operacionalmente definida no es suficiente es cierto. Como lo dice Hacking (1983), podemos medir el CI y fanfarronear porque una docena de técnicas diferentes nos dan el mismo patrón de números, aún cuando no tenemos ni la más mínima comprensión causal del fenómeno. Pero identificar una estructura subyacente en el fenómeno tampoco es una cuestión sencilla. Toomela (2008) hace una crítica a la psicología cuantitativa en la que ofrece varias razones por las cuales uno debería detenerse en el análisis de cómo se generan teorías a partir del uso de la herramienta estadística. Argumenta por ejemplo que es posible encontrar regularidades en el comportamiento que no provienen de una estructura o causa mental oculta, un fenómeno de tal tipo puede meramente ser creado, por ejemplo, cuando los grupos de datos son analizados. También

argumenta que eventos similares externamente pueden emerger de diferentes mecanismos internos y que las variables codifican información sólo sobre eventos, en principio. Un tercer argumento sostiene que las variables pueden contener información sobre eventos en diferentes niveles de análisis, y muchos elementos pueden actuar colectivamente en una situación, de modo que su comportamiento puede ser caracterizado por uno o algunos pocos *parámetros de orden* o *variables colectivas* (Smith y Thelen 1993). El problema, en general, es que las variables no representan directamente los fenómenos mentales sino sólo el comportamiento. La psicología, en su opinión, debería contar con estudios cualitativos que vayan más allá de la observación de relaciones cuantitativas entre variables. Los niveles de análisis cualitativo deberían ser claramente distinguidos y debería tomarse en cuenta que los elementos no son siempre independientes en principio. Toomela luego opta por la caracterización tipológica y urge por una teoría sustantiva para guiar la selección de los datos. Está criticando al análisis estadístico y por ello vela por la interpretación del significado de los datos, sin embargo creo que aquí se aleja de la problemática de la medición y que el tipo de interpretación más interesante es la que podamos dar desde el punto de vista del uso de los instrumentos. Son los usos de la individuación causal y de la correlación estadística los que deben ser puestos en contexto y ponderados adecuadamente.

Boumans en 2004 sostiene que el problema principal es que la teoría representacional de la medición se ha tornado una disciplina puramente matemática, dejando de lado la cuestión acerca de cómo las estructuras matemáticas ganan significado empírico en la medición, que conlleva procedimientos y dispositivos concretos de medición. Este punto de vista concurre con la importancia del estudio de la nomología que guarda el uso de los instrumentos, o los instrumentos mismos, del estudio del modo en que se construye un instrumento y los supuestos para la recolección de datos así como del híbrido fenoménico con el que un instrumento lidia, como preocupaciones esenciales para el estudio de la medición.

Lo que remarco aquí es la importancia del estudio de casos. Por ejemplo para entender en qué punto la medición de un fenómeno descansa en modelos que acaban por proponer un concepto, en lugar de sondear su comportamiento de manera que sea conmensurable con datos reales. En estos casos hay que diferenciar el modo de construcción de un instrumento de medición del de modelos matemáticos. Cobra importancia además estudiar el uso que se hace de las variables con relación al tipo de fenómeno que se trata y cuáles son las variables con medición confiable que se le asocian. Es de interés investigar cómo las prácticas hablan causalmente, es decir cómo se puede distinguir por dónde pasa la fuerza nomológica de un instrumento, teniendo en cuenta que las entidades del caso son difíciles de medir y aparecen en contextos multicausales, no lineales, etc. Otro punto de

importancia consiste en saber que no siempre se tiene una teoría por delante aunque sí conceptos, y que estos conceptos no requieren ser siempre cuantificados para obtener con ellos más precisión sobre el fenómeno. Un último llamado de atención debería ir dirigido hacia la construcción de modelos híbridos y la dinámica entre modelos matemáticos, datos de medición y fenómenos eminentemente cualitativos.

Comentarios finales

No ha sido aquí el lugar de revisar casos pero analizando la historia reciente de la psicología es fácil reconocer el tipo de problemas y construcciones arriba referidos (Smith y Thelen 1993, van Geert 1994). El último punto en el que me quiero detener hace alusión a las dificultades que tiene plantear la idea de un sesgo en el representacionalismo que por supuesto es también una posición con grises de interés. Michell (2003) defiende que el positivismo, léase el representacionalismo, no ha sido ni es una barrera para los métodos cualitativos. Éste es un sesgo de tradición pitagórica, acerca de que los atributos son fundamentalmente cuantitativos, que renace en cualquier modo de realismo ingenuo pero que no es característico de los diferentes modos de kantismo e idealismo que pueden identificarse también en el representacionalismo. Michell relata cómo Okham fue capaz de ver a la matemática más como un instrumento, como un lenguaje aplicable en muchas disciplinas e interpretable de muchas maneras. Destaca también cómo estructuras no cuantitativas como ciertas teorías sobre lenguajes formales han tenido aplicaciones productivas. Sin embargo, el contrapunto que he pretendido marcar se refiere al tratamiento matemático de precisión, de interés para la teoría de la medición inspirada en la metrología, que desarrollan ciertos modelos en psicología. Esto es, el interés que pueda tener el estudio de los instrumentos en el contexto del estudio de fenómenos particulares, más allá de la posibilidad de describirlos como estructuras lógicas.

REFERENCIAS

Boumans, M.J. 2005, 'Measurement Outside the Laboratory', *Philosophy of Science* 72(5), 850-863.
---------- 2004, 'The reliability of an instrument', *Social Epistemology* 18, 215-246.
---------- 2001, 'Fisher's instrumental approach to index numbers', en Klein, J., Morgan, M. (Eds.) *The Age of Economic Measurement*. Annual Supplement to *History of Political Economy*, 33, Duke University Press, Durham, 313–344.
Carnap, R. 1950, *Logical Foundations of Probability*, University of Chicago Press.
Cartwright, N. & Chang, H. 2008, 'Measurement', *The Routledge Companion to Philosophy of Science*, Psillos, S. & Curd, M. (eds.), New York, Routledge, 367-375.
Díez, J. 1997a, 'A Hundred Years of Numbers. An Historical Introduction to Measurement Theory. Part I: The Formation Period', *Studies in History and Philosophy of Science* 21(1), 167-181.

Díez, J. 1997b, 'A Hundred Years of Numbers. An Historical Introduction to Measurement Theory. Part II: Suppes and the Mature Theory', *Studies in History and Philosophy of Science* 22(2), 237-265.

Diez, J. & Moulines, C. 1997, *Fundamentos de filosofía de la ciencia*, Ariel.

Finkelstein, L. 1982, 'Theory and philosophy of measurement', en Sydenham, P. (Ed.), *Handbook of Measurement Science vol. 1: Theoretical Fundamentals*, Wiley, New York.

Hacking, I. 1983, *Representing and Intervenig, Introductory Topics in the Philosophy of Natural Science,* Cambridge University Press.

Helmholz, H. 1921, *Epistemological Writings*, Dordrecht, Reidel (1977).

Hempel, C. 1952, *Fundamentals of Concept Formation in Empirical Sciences,* University of Chicago Press

Michell, J. 1999, *Measurement in Psychology. Critical history of a methodological concept.* Cambridge University Press.

---------- 2003, 'The Quantitative Imperative: Positivism, Naive Realism and the Place of Qualitative Methods in Psychology', *Theory Psychology*, 13, 5-31.

Ruspini, E. 2002, *Introduction to Longitudinal Research in Social Sciences*, Routledge.

Smith, L. & Thelen, E. (Eds.) 1993, *A Dynamic Systems Approach to Development*. Cambridge, MA: MIT Press.

Stevens, S. 1946, 'On the Theory of Scales of Measurement', Science 103 (2684), 677–680.

Toomela, A. 2008, 'Variables in Psychology: A Critique of Quantitative Psychology', Integrative Psychological and Behavioral Science, 42(3), 245-65.

Van Geert, P. 1994, *Dynamic systems of development. Change between complexity and chaos.* Harvester.

THEORETICAL PREDUCTION AND COMPUTATIONAL SCIENTIFIC DISCOVERY

ANDRÉS RIVADULLA[1]

ABSTRACT: Can deductive reasoning be applied to the context of discovery of mathematical physics? I term *preduction* the method of reasoning that, starting from the available theoretical background as a whole, allows for the anticipation of previously unknown results, provided that the combination of already accepted results of different theories – those taken as premises of the deductive reasoning – is compatible with dimensional analysis. This is the method by which many hypotheses, laws, and theoretical models are introduced into physics. C*omputational preduction* further extends the possibilities of machine learning, which in the past forty years has informed computational systems, implementing both the rediscovery of empirical laws and the automated discovery of equations in data bases. Automated preduction would also facilitate scientific creativity in theoretical physics.
KEY WORDS: scientific discovery, theoretical preduction, computational discovery, computational preduction

1. Problem solving and induction

Two years after the publication of Popper's *Objective Knowledge*, a book in which he develops his anti-inductivist stance, which states that we can never rationally justify our belief in the truth of a theory, or in its being probably true, Herbert Simon (1974, 330) proclaimed the close relationship between induction and problem solving: "from a logical standpoint the processes involved in problem solving are inductive, not deductive." More specifically: "The problem solving process is an information gathering process as much as it is a search process. ... The process for using this information to steer the search are generally processes of inductive inference. Being inductive, they do not provide certainty, but have only heuristic value in guiding the search and making it efficient."

Since both positions are not compatible with each other, one possible outcome is to assume that the logical illegitimacy of induction is not decisive for the methodology of science. Langley *et al.* (1987, 16-17.My italics, A. R.) are perfectly aware of the weakness of inductive inferences. In an argument reminiscent of Popper, they maintain that:

> "It is well known that no universally quantified generalization can be verified decisively by a finite set of observations. Not only may subsequent observations overthrow the verification, but it will never be the case that the generalization that

[1] Projects FFI2009-10249 and AIB2010PT-00106 supported by the Spanish Science and Innovation Ministry.

> has been found to fit the observations is unique. Other generalizations, known or not, will describe or explain the data just as well as the one being considered."

But, instead of concluding, with Popper, that induction simply does not exist, they conclude that *science solves these problems of certainty and uniqueness by ignoring them*. In relation to the *flawed assumption* that we are seeking an *infallible* method of scientific discovery, they affirm:

> "The discovery process solves the problem of infallibility, as it does the problems of certainty and uniqueness of generalizations, by ignoring it. ... We require neither a guarantee that solutions, in the form of laws, can be found, nor a guarantee that the laws can be strongly verified." (*op. cit.*, 303)

Insisting on this idea, Langley *et al.* (*op. cit.*, 16-17.My italics, A. R.) maintain that

> "The problem of discovery is precisely to find some generalization that is parsimonious and consistent with all the given data (within some criterion of accuracy). At the point of discovery, no claim is made that the generalization will survive the test of additional evidence, or that another generalization will not be found that is equally adequate for describing the original data. (...) If even one generalization or explanation can be found to fit a body of data reasonable well, it will usually be accepted, at least until a competitive law or explanation comes along or until new important data are found that do not fit at all. When two or more explanations for the same data or for overlapping data are discovered, there arise the new tasks of evaluating them relative to one another and (usually) discovering new laws that reconcile the discrepancies. In all these circumstances, the first and vital task is to discover laws. While that is going on, the existence of competitors, actual or potential, is simply irrelevant. *The concern that the philosophy of science has had for the certainty and uniqueness of inductions stems from a misconception of the nature of science*. … Laws, or supposed laws, will often be discovered (a happy moment). They will often be refuted by later evidence (a sad moment), and they will sometimes have to contend with rival laws that explain the same data (a perplexing moment)."

Thus Langley's *et al.,* (*op. cit., ibid.*) position on the Humean *induction problem* is:

> "The function of verification procedures is not to provide scientists with unattainable certainty or uniqueness for discoveries, but to inform them about the risks they are running in committing themselves to hypotheses that have been discovered and to provide guidance that may enhance their chances of making relatively durable discoveries."

Popper's issues with the induction problem do not seem to be justified.

Moreover Langley *et al.* (*op. cit*., 39 and 40, respectively) are very far from endorsing a naive verificationism:

> "The notion that theories are to be validated by deriving them deductively from observations – usually called *naive verificationism* – is clearly invalid." "Of course, many scientific papers contain statements about evidence 'supporting', 'strengthening', or 'weakening' a theory. ... But such statements fall far short of constituting quantitative claims of 'degree of justifiable belief'. There does not seem to be any essential requirement in science for such claims."

As to the relationship between the context of justification and the context of discovery, which for them are not separate and distinct questions (see Rivadulla 2012), Langley *et al.* (*op. cit*., 58-59) conclude:

> "First, a new discovery … is generally developed through a step-by-step search guided at each step by data. Second, at each step in this process, the developing hypothesis is guaranteed to be confirmed by the data examined thus far. Third, since only a finite body of data is in question, the verification can be complete (that is, compatibility with the data can be definitely confirmed), although uniqueness cannot be guaranteed. Fourth, because of the incremental character of the process, the discovery, once arrived at, has already achieved a considerable degree of confirmation, although it still may fail when tested against new data."

Paul Thagard (1988, 27-28) also points to the limitations of induction:

> "inductive inference [is] inference to conclusions that do not follow necessarily from what is already known. Induction … can lead to conclusions that are not only useless but false. At least in deduction, what follows from true premises has to be true, whereas induction unavoidably involves a kind of leap. You may have observed myriad instances of copper conducting electricity, but generalizing that all copper conducts electricity still introduces uncertainty."

Thus, according to Thagard, the use of induction cannot go beyond the production of rules for problem solving and explanation:

> "in making inductions, we do not want to be concerned only with reaching well-warranted conclusions: we want to produce rules that will be useful for future use in problem solving and explanation."

This position is in complete agreement with Simon's (1995, 175. My italics, A. R.):

> "Nor is exact truth the realistic goal of discovery. … The task of discovery is to create theories that explain the data 'well enough', with the hope that tomorrow they can be explained better. … We call a problem solution a discovery to the extent that it is both valuable (and/or interesting) and novel. So *an empirical theory of discovery is a theory of how novel and valuable* (*or interesting*) *problem solutions are in fact found*."

2. Automated induction and computational scientific discovery

To claim that scientific discovery proceeds computationally points to the existence of methods, systems or programs that allow for its automation. As Saso Dzeroski *et al.* (2007, 1) affirm: "Research on computational scientific discovery aims to develop computer systems which produce results that, if a human scientist did the same, we would refer to as discoveries."

The computational science of scientific discovery is grounded in the development of computational systems that from their very beginning have been able to reproduce, replicate, simulate, or rediscover important laws of the history of science. Such systems or programs, usually written in LISP or in PROLOG, did appear during the 1960s, and since the 70s they have become well established. From the perspective of the twenty-first century, referring to the set of programs collectively called BACON, Dzeroski *et al.* (2007, 8.My italics, A.R.) affirm that they implemented

> "data-driven induction of descriptive laws and were demonstrated on historical examples. *Together, they provided the first compelling evidence that computational scientific discovery was actually possible.*" "The early work on computational discovery focussed on reconstructions from the history of science that were consistent with widely accepted theories of human cognition. This was an appropriate strategy, in that these examples let researchers test their methods on relatively simple problems for which answers were known." (*op. Cit.*, 9)

Langley's *et al.* (1987, 20.My italics, A. R.) book is mainly concerned with "*the induction of descriptive and explanatory theories from data.*" In order to implement this task, these authors employ the computer program BACON, which is capable of inductively producing scientific discoveries from empirical data. Referring to BACON, Shrager and Langley (1990, 10) claim: "BACON … focused on the induction of numerical laws from experimental data." And Thagard & Nowak (1990, 28) affirm that "the BACON program takes numerical values as inputs and generates equations that fit those values." Also, Herbert Simon (1992, 8.My italics, A. R.) insists on the pure inductive character of BACON:

> "*A series of computer programs, collectively named BACON, has successfully simulated the process whereby a substantial number of important laws of eighteenth and nineteenth century chemistry were induced from data.*"

For instance, BACON is able to obtain Kepler's Third Law inductively from data. But it can also obtain Joseph Black's laws of temperature equilibrium, the law of conservation of momentum, and many others. It does not only find laws, but is also able to introduce new theoretical concepts, like inertial mass, specific heat, atomic weight, etc. Simon (1995, 177) also recognizes BACON's capabilities as a data-driven automated discovery program:

> "It can be claimed that BACON's process resembles closely the process that human scientists have used in a number of important historical cases where little or no theory was available to guide the search for laws: hence 'data-driven discovery'."

3. Computational preduction

Computational scientists repeatedly assert that they have filled the gap left behind by philosophers of science, and they reproach them for abandoning the context of discovery (see Rivadulla 2012). Is the moment now ripe for the philosophers of science to regain the lost ground and to advance new possibilities in computational science? In order to answer this question, it is necessary to extend the spectrum of possibilities of scientific discovery. In this respect, both a wider concept of 'scientific discovery' and a closer attention to real scientific methods are unavoidable. In particular we must investigate the possibility of new strategies for the introduction of novel ideas into science, beyond standard ampliative inferences, such as induction and abduction.

As to the first issue, Dzeroski *et al.* (2007, 3-4.My italics, A. R.) define scientific discovery – briefly, and without further reference to exclusive methods of research – as

> "the process by which a scientist creates or finds some hitherto unknown knowledge, such as a class of objects, an empirical law, or an explanatory theory. … A defining aspect of discovery is that the knowledge should be *new and previously unknown*."

When addressing the issue of non-standard ways of introducing new ideas into science, philosophers of science have maintained different perspectives. Peter Medawar (1974, 289) pointed to the weakness of deductive reasoning in scientific methodology:

> "The weakness of the hypothetico-deductive system, insofar as it might profess to offer a complete account of the scientific process, lies in its disclaiming any power to explain how hypotheses come into being."

For his part, Thomas Nickles (2008, 446) guesses at the applicability of deductive reasoning in discovery tasks:

> "it is worth noting that even an ordinary deductive argument need not be sterile: it may be epistemologically ampliative even though it is not logically ampliative, for we are not logically omniscient beings who see all the logical consequences of a set of propositions."

And Elie Zahar (1983, 244-245.My italics, A. R.) has unambiguously advanced the thesis of the deductive character of scientific discovery:

> "*the process of discovery … rests largely on deductive arguments from principles which underlie* not only *science* and deductive metaphysics but also everyday decisions."

On pages 249-250 he argues that

> "the logic of scientific discovery is not inductive; neither does it resemble artistic creation, but is, instead, largely deductive. … Unlike Popper, I think that deduction constitutes the most important element in the process of invention."

The problem with Zahar's proposal is that the deductive character of scientific discovery is based on the metaphysical assumption of the existence of several *philosophical principles* and *meta-principles*, which, as a sort of premises for deductive reasoning, allow for the creation of theories. Such principles are the *Principle of Correspondence,* the *Principle of Identity,* the *Principle of Sufficient Reason,* and the *Principle of Proportionality of Cause and Effect*. Interesting though it is, Zahar's proposal is a metaphysical theory about the deductive nature of scientific discovery. It is a highly speculative hypothesis, far removed from the methodological character both of ampliative inferences in natural observational sciences and of anticipative inferences, like theoretical preduction, in natural theoretical sciences such as mathematical physics.

Contrary to philosophers of science, computational scientists seem to be less reticent when debating the role of deductive reasoning in scientific discovery. For instance, Shrager and Langley (1990, 5-9) emphasize the *deductive law formation*, which

> "produces laws by a second route [the first one is the inductive one, A. R.], starting with a theory and using an explanatory framework to deduce both a law and an explanation of how that law derives from the theory. … For instance, Einstein's theory of general relativity led to an inferred law about the orbit of Mercury."

Indeed, this is a clear instance of how a novel law can be derived in a purely *intra-theorical* way. Further examples are the deduction of Einstein's famous law $E=mc^2$, Maxwell-Boltzmann distribution laws, etc.

The possibility of deductive discovery brings the philosopher of science closer to the idea of *preduction*. As I have advanced elsewhere (see Rivadulla 2010), preduction anticipates still-unknown ideas in physics by the combination, compatible with dimensional analysis, of previously accepted results of different disciplines or theories of mathematical physics. Since these results, which are the premises of the preductive way of reasoning, proceed from different theories, preduction represents *transverse* or *inter-theoretical* deduction. This is what makes it possible to anticipate new ideas in physics.

Is *theory-driven discovery* an antecedent of computational preduction? Simon (1992, 10-11.My italics, A. R.) offers several examples of computational discovery:

> "Suppose we provide BACON with some very broad theoretical concepts: that when substances are mixed together both mass and heat are conserved; and that the law

> describing the temperature equilibrium of such a mixture should be symmetrical in the properties of the components. Then, Black's law of temperature equilibrium can be deduced from these assumptions, in advance of any examination of data, and the data simply used to confirm the law. Without these assumptions, BACON must induce the law from data with the help of rather arduous calculations. Thus, assumptions of conservation and symmetry can be used as heuristics to reduce the search required to find laws."

Similarly, under the assumption that chemical substances are made up of molecules, and molecules of atoms of the elements, and that atoms are conserved in reactions and that volumes of gasses, under constant temperature, are proportional to the numbers of their molecules, etc., the DALTON program

> "deduces the chemical formulas of the molecules involved. For example, on being told that three volumes of hydrogen and one of nitrogen produce two of ammonia, it concludes, correctly, that hydrogen and nitrogen are H_2 and N_2, respectively, and that ammonia has the formula NH_3." (Simon 1992, *ibid.*)

Thus, insofar as theory-driven discovery might involve assumptions from different theories, it could conceivably advance computational preduction.

Computational preduction is automated theoretical preduction. The *computational preduction* hypothesis assumes the possibility of automating preduction mechanisms. It is grounded in the existence, since the 1970s, of computational methods for automating scientific discoveries. Since the applicability of the methods of IA to scientific discovery allows for the automation of inductive and abductive processes, computational preduction would help us to *anticipate* new and previously unknown results in theoretical physics, and would therefore extend scientific creativity.

REFERENCES

Dzeroski, S., *et al.,* 2007, 'Computational Discovery of Scientific Knowledge', in Dzeroski, S. and Todorovski, L. (eds.), *Computational Discovery of Scientific Knowledge. Introduction, Techniques, and Applications in Environmental and Life Sciences*. LNAI 4660, Berlin-Heidelberg, Springer

Langley, P. *et al.,* 1987, *Scientific Discovery. Computational Explorations of the Creative Processes*, Cambridge, MA, The MIT Press

Medawar, P. 1974, 'Hypotheses and Imagination', in Schilpp, A. (ed.), *The Philosophy of Karl Popper*, La Salle, Ill., Open Court, 274-291

Nickles, T., 2008, 'Scientific Discovery', in Psillos, S. & Curd, M. (eds.), *The Routledge Companion to Philosophy of Science*. London, Routledge

Rivadulla, A. 2010, 'Complementary Strategies in Scientific Discovery: Abduction and Preduction', in Bergman, M. *et al.* (eds.), *Ideas in Action*: *Proceedings of the Applying Peirce Conference*, Nordic Studies in Pragmatism 1, Nordic Pragmatism Network, Helsinki, 264-276

Rivadulla, A. 2012, 'Ciencia computacional de la ciencia. Filosofía del descubrimiento automático', in Fernández Moreno, L. *et al.* (eds.), *Ensayos sobre lógica, lenguaje, mente y ciencia*. Sevilla, Alfar

Shrager, J. and Langley, P., 1990, 'Computational Approaches to Scientific Discovery', in Shrager and Langley, *Computational Models of Scientific Discovery and Theory Decision*, San Mateo, California, Morgan Kaufmann Publishers

Simon, H. A. and Lea, G., 1974, 'Problem Solving and Rule Induction', in Simon, H., *Models of Thought*, New Haven and London, Yale University Press 1979

Simon, H. A., 1992, 'Scientific Discovery as Problem Solving', *International Sudies in the Philosophy of Science* 6, No. 1, 3-14

Simon, H. A., 1995, 'Machine Discovery', *Foundations of Science* 2, 171-200. Reprinted in Zytkow, J. (ed.), *Machine Discovery*, Kluwer Academic Press, Dordrecht 1997

Thagard, P., 1988, *Computational Philosophy of Science*, Cambridge, MA, The MIT Press

Thagard, P. and Nowak, G., 1990, 'The Conceptual Structure of the Geological Revolution', in Shrager and Langley, *op. cit.*

Zahar, E., 1983, 'Logic of Discovery or Psychology of Invention?', *Brit. J. Phil. Sci*. 34, 243-261.

LA PROBLEMÁTICA NATURALIZACIÓN DE LA NORMATIVIDAD NATURAL. ORGANIZACIÓN Y MALFUNCIONALIDAD BIOLÓGICA

CRISTIAN SABORIDO
UPV
UNED

ABSTRACT: In this paper, we analyze the way in which mainstreams theories on biological functions, i.e. the evolutive-etiological and the classical systemic-dispositional approaches, address the notions of *natural normativity* and *biological malfunction*. Then, we show why these approaches do not account for the specific normative dimension of malfunctions. Finally, we expose a new organizational approach of biological normativity that can be developed in order to ground adequately the concept of malfunction and supposes a promising way for the naturalization of biological norms.
KEYWORDS: Normativity, function, naturalism, disease.

> "Decir que los riñones están fallando es fácil. Eso lo ve cualquier idiota. Lo interesante es poder explicar *qué significa eso.*"
>
> Dr. House

1. Naturalismo y normatividad

La posibilidad de una *normatividad natural* es una cuestión que ha sido abordada desde distintas perspectivas por parte de la Filosofía de la Ciencia (Davies 2001, Barham 2012, Christensen 2012). Se trata, de hecho, de una cuestión muy importante para la misma perspectiva naturalista. Hume (1978) ya argumentó que es erróneo extraer conclusiones normativas de premisas descriptivas, pues esto nos aboca necesariamente a la "falacia naturalista", en la cual pasamos del "ser" al "deber ser" sin justificación. Así, concluir de una descripción del comportamiento de los rasgos biológicos que estos "deben" comportarse de tal manera es una argumentación falaz. Del hecho, por ejemplo, de que *de facto* el corazón de los vertebrados bombee sangre, no podemos inferir que el corazón *deba* bombear sangre. Y algo parecido defendió también Moore (1903), al

afirmar que una propiedad normativa (como "bueno" o "malo") y una propiedad natural (por ejemplo, "placer" o "bienestar") no pueden ser nunca conceptualmente idénticas, por lo que cualquier intento de identificar una con la otra será siempre lo que Moore denomina una "cuestión abierta" ("*open question*"). No es filosóficamente aceptable, por tanto, identificar enunciados como "contribución a la reproducción y la supervivencia" o "favorecido por la acción de la selección natural" con proposiciones normativas tales como "comportamiento funcional correcto".

2. Normas y funciones biológicas

Recientemente, autores como Cummins (2002), Davies (2009) o Fraansen (2009) han recuperado estos planteamientos en el marco de la discusión filosófica acerca del concepto de función biológica. Estos autores defienden vehementemente que un análisis teórico del concepto de función biológica no puede dar cuenta de una presunta dimensión normativa. Estos autores sostienen, contra la extendida interpretación de que la noción de función es inherentemente normativa (Bedau 1992, Neander 1995), que no está teóricamente justificado extraer una atribución normativa de una atribución funcional. De hecho, el creer que las atribuciones funcionales hacen referencia a normas que subyacen al comportamiento de los seres vivos es únicamente una consecuencia de un "conservadurismo conceptual" que proviene de nuestra herencia cultural y de nuestros esquemas psicológicos y que nos lleva a interpretar el mundo natural como producto de un diseño intencional consciente.

No obstante estos argumentos, el comportamiento de los sistemas biológicos es tradicionalmente entendido de una forma normativa. En palabras de Canguilhem:

> "La vida es, de hecho, una actividad normativa. Lo normativo, en filosofía, incluye todo juicio que evalúa o califica un hecho en relación a una norma, pero este tipo de juicio está esencialmente subordinado a aquello que establece la norma. Lo normativo, en el sentido completo del término, es aquello que establece normas. Es en este sentido en el que hablamos de normatividad biológica." (Canguilhem 1966, 126)

Muchos análisis teóricos del concepto de función biológica siguen esta línea de pensamiento y defienden que describir un rasgo de un ser vivo como funcional implica hacer una distinción normativa. Atribuir una función determinada a un rasgo biológico significa atribuir a este rasgo una norma a la cual su actividad *debe* someterse. La noción de funcionalidad biológica, según defienden muchos autores, está vinculada a la de normatividad natural (McLaughlin 2009, Price 1995). Una vez que afirmamos que la función natural de X es F, estamos diciendo que producir F es una norma propia de X. De este modo, se considera que las funciones biológicas son normativas, al igual que

se considera normalmente que ocurre en el caso de las funciones de los sistemas socio-técnicos. Así, muchos teóricos defienden que el corazón *debe* bombear sangre, del mismo modo que usualmente se afirma que los faros de un coche *deben* alumbrar la calzada o que los bomberos *deben* sofocar los incendios.

Sin embargo, la adscripción de normas a las entidades naturales es un reto complicado e inevitable para toda teoría naturalista sobre las funciones y la normatividad. Las normas naturales parecen precisar de un tipo de fundamentación teórica diferente al de las normas de las entidades y sistemas socio-técnicos. Dado que en el caso biológico ya no podemos contar con un diseño intencional que nos sirva como referencia, ¿existe un "modelo" que podamos usar como criterio para evaluar los comportamientos funcionales de los rasgos de los seres vivos? Atribuir normas a un rasgo concreto implica hacer una referencia a algún efecto específico que creemos de especial relevancia para este rasgo, lo cual permite que su actividad actual pueda ser normativamente evaluada (Neander 1995, McLaughlin 2009, Krohs 2010). Recientemente, Tyler Burge ha defendido que, en un sentido básico, una "norma natural" es:

> "un nivel de realización de una actividad que es adecuado para el cumplimiento de una función o propósito que es explicativamente relevante independientemente de cualquier consideración positiva o negativa de un observador externo" (Burge 2009, 269).

Según esta interpretación, una norma natural se fundamenta en el cumplimiento de una función biológica y esto supone que una fundamentación adecuada de la normatividad biológica debe ser capaz de explicar por qué consideramos un efecto y no otro como la función o el propósito que el rasgo *debe* llevar a cabo independientemente de ningún criterio o interés de un observador externo. Dicho en otras palabras, una teoría que entienda las funciones como normativas y las normas como naturales ha de explicar cómo se fundamenta, en términos naturalistas, esta normatividad.

3. El concepto normativo de *malfunción*

Un importante concepto normativo derivado del concepto de función es el de *malfuncionalidad*. Aunque este concepto es muy utilizado tanto en contextos habituales como en las ciencias biológicas, su fundamento teórico ha recibido muy poca atención por parte de la Filosofía de la Biología como una cuestión independiente. Y esta no es una cuestión trivial: el discurso funcional parece suponer la posibilidad de hablar de *disfunciones* y *malfunciones.*

Así, una de las primeras cosas que hay que tener en cuenta al abordar esta cuestión es que *el concepto de malfuncionalidad es diferente del concepto de no-funcionalidad*, con el que se suele

confundir a menudo. La normatividad descriptiva ligada al concepto de función se da en términos de "todo o nada", una función se cumple o no se cumple y, por tanto, un rasgo cumple totalmente con la norma o no cumple con ella en absoluto. Por el contrario, la malfuncionalidad es un concepto gradual que permite hablar de todo un rango de posibilidades en el cumplimiento de una norma natural: una malfunción hace referencia a una norma que puede cumplirse mejor o peor. La malfuncionalidad es una cuestión de grado (Krohs 2010, 342). Una malfunción puede ser algo letal o simplemente un inconveniente leve. Un rasgo es malfuncional cuando lleva a cabo su función de una forma incorrecta y esto puede hacerlo con diferentes grados de relevancia para el sistema.

Defenderemos por lo tanto que l*as normas que fundamentan el concepto de malfunción y las normas que fundamentan el concepto de función normativa son de un tipo diferente.* Una caracterización teórica del concepto de malfunción precisa de una fundamentación de un tipo de normatividad más sofisticado. Más allá de determinar cuáles son las normas a las que el funcionamiento de un sistema obedece *de facto*, una teoría sobre las malfunciones debe ofrecer los recursos teóricos necesarios para poder decir cuándo estas normas se están llevando a cabo de un modo incorrecto, y en qué medida. Así, es posible distinguir dos tipos de normatividad diferentes: una normatividad relacionada con la idea de función, y otra normatividad ligada a la malfuncionalidad. Una función puede ser llevada a cabo correcta o defectuosamente, y esto implica postular que existe una *norma* que puede o no ser cumplida y que nos permite evaluar en términos de "bueno" o "malo" un comportamiento funcional. Por tanto, una fundamentación adecuada y completa de la normatividad biológica debe ser capaz de dar cuenta de las malfunciones y de sus normas.

Cabe señalar que esto tiene importantes repercusiones para la Filosofía de la Medicina. El concepto de malfunción ha sido utilizado para justificar la distinción conceptual entre salud y enfermedad y algunos de los teóricos más influyentes (Boorse 1977, Wakefield 1992, Schramme 2007) consideran que esta noción de malfunción es un ingrediente necesario para una adecuada definición de enfermedad. Partiendo de distintas nociones de malfunción, se ha vinculado la noción de enfermedad (fisiológica o mental) con la presencia de un funcionamiento incorrecto de algún rasgo biológico. Así, Boorse define enfermedad como "un tipo de estado interno que impide al organismo mantenerse sano, esto es, que reduce al menos una de sus habilidades funcionales por debajo de su eficiencia típica" (Boorse 1977, 62), y Wakefield afirma que una condición física constituye una enfermedad si esta condición "es resultado de una incapacidad de algún mecanismo interno para realizar su función natural al nivel adecuado" (Wakefield 1992, 384). Aún cuando se argumente -tal y como tradicionalmente han hecho los teóricos "normativistas" (Nordenfelt 1987,

Reznek 1987)- que el concepto de malfunción no basta para dar cuenta de la distinción salud/enfermedad, pues esta distinción depende íntimamente de todo un conjunto de valores, creencias y percepciones subjetivas, el concepto de malfunción biológica es de una importancia indiscutible para la Filosofía de la Medicina contemporánea.

4. Las perspectivas etiológica y disposicional de las malfunciones

Dentro de los principales enfoques en el debate sobre las funciones biológicas se han propuesto diferentes justificaciones teóricas de la noción de función normativa. Así, para los partidarios del enfoque etiológico-evolutivo (Wright 1973, Millikan 1989, Godfrey-Smith 1994) la norma de las funciones se fundamenta en las condiciones evolutivas de existencia del rasgo funcional. Por otro lado, dentro del enfoque disposicional (Cummins 1975, Davies 2001, Craver 2001) se defiende que un rasgo cumple con su norma cuando contribuye al logro de una meta o finalidad del sistema. A pesar de sus diferentes estrategias, ambos enfoques adolecen de un importante problema con respecto a su interpretación de la dimensión normativa de las funciones, pues se muestran incapaces de fundamentar la dimensión normativa que atañe al concepto de *malfunción*.

Aún cuando pudiera defenderse que las propuestas etiológicas o disposicionales son capaces de ofrecer teorías válidas para una caracterización de la noción de función normativa, son insuficientes para ofrecer una definición adecuada del concepto de malfunción biológica. El enfoque etiológico-evolutivo afirma que un efecto de un rasgo de un organismo pertenece a un tipo funcional cuando este rasgo ha sido heredado de una instancia ancestral de ese organismo que poseía asimismo instancias ancestrales de ese rasgo que han sido seleccionadas evolutivamente precisamente por poseer este mismo efecto. Así pues, solamente si un rasgo de un tipo funcional específico posee este efecto evolutivamente significativo, podemos decir que este rasgo es funcional. El problema es que para ser capaz de justificar una normatividad evaluativa que nos ofrezca los recursos teóricos necesarios para poder afirmar que algo está "funcionando correctamente" o "funcionando defectuosamente" una teoría de las funciones requiere de algo más que de una falta de "función seleccionada". Los teóricos etiologicistas defienden que su enfoque puede fundamentar el concepto de malfunción pero, en realidad, están confundiendo la malfuncionalidad con la no-funcionalidad (cfr. Davies 2001).

Por otro lado, el enfoque sistémico-disposicional es vulnerable a la misma crítica que se aplica a los enfoques etiológico-evolutivos**.** Si interpretamos una función como una contribución o disposición a lograr una meta sistémica, sólo podemos considerar que un rasgo es funcional cuando

la contribución o disposición está presente y que un rasgo es no-funcional cuando esta contribución o disposición está ausente. Desde este tipo de definición solo es por lo tanto posible hablar en términos de todo o nada, esto es, de funcionalidad o no funcionalidad. Como en el caso de la perspectiva etiológica, el enfoque sistémico no es capaz de dar cuenta del concepto gradual de malfunción.

5. Hacía una noción organizacional de la malfuncionalidad biológica

Probablemente, la mejor forma de lograr un análisis teórico adecuado del concepto de malfunción sea partiendo del nuevo *marco organizacional* del debate sobre las funciones en Filosofía de la Biología (Schlosser 1998, Christensen & Bickhard 2002). Esta perspectiva organizacional -y en concreto una propuesta que interpreta las funciones como contribuciones al auto-mantenimiento biológico- puede ser desarrollada a fin de fundamentar adecuadamente el concepto de malfunción. El argumento central de esta propuesta organizacional es que una función es un efecto específico (diferenciado) de un rasgo que supone una contribución al mantenimiento de un cierre causal que se da entre distintos efectos de las estructuras y procesos de un sistema. Este efecto funcional es una condición necesaria para el auto-mantenimiento de una organización biológica que produce y mantiene al rasgo y cuya existencia depende a su vez de la acción diferenciada de sus componentes (Mossio *et al*. 2009, Saborido *et al.* 2010, 2011).

Si aplicamos este modelo teórico al caso de los rasgo malfuncionales debemos decir que un rasgo que malfunciona es, ante todo, un rasgo funcional, pues contribuye al auto-mantenimiento de la organización, pero es un rasgo que no realiza esta contribución de acuerdo con ciertas normas que se fundamentan en las propiedades de la misma organización del organismo. Es por esto que podemos decir que un comportamiento funcional supone una mala o defectuosa contribución. Los rasgos biológicos bajo esta interpretación muestran una gradualidad en su eficiencia funcional. Una malfunción no es un comportamiento en términos de todo o nada, como es el caso de la distinción función / no-función, sino que los efectos de un rasgo funcional pueden ser calificados como "buenos" o "malos" atendiendo a una norma corporeizada (*embodied norm*) que se basa en la acción del subsistema regulatorio que el resto del organismo presupone y que permite todo un rango de posibles comportamientos malfuncionales.

De este modo, partiendo de un enfoque de raíces aristotélico-kantianas, las propiedades de *presuposición funcional* y de *regulación adaptativa* que muestran los sistemas biológicos nos permiten salvar las limitaciones de los diferentes análisis teóricos que se han presentado sobre el concepto de función, ofreciéndonos los recursos teóricos necesarios para desarrollar un enfoque

organizacional capaz de justificar teóricamente la atribución de malfunciones biológicas y, por lo tanto, abriendo una prometedora vía para la naturalización de normas naturales.

REFERENCIAS

Barham, J. 2012, 'Normativity, Agency, and Life', *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, 92-103.

Bedau, M.A. 1992, 'Goal-Directed Systems and the Good', *The Monist* 75, 34- 49.

Boorse, C. 1977, 'Health as a Theoretical Concept', *Philosophy of Science* 44(4), 542-573.

Burge, T. 2009, 'Primitive Agency and Natural Norms', P*hilosophy and Phenomenological Research* 89 (2), 251-278.

Canguilhem, G. 1966, *Le normal et le pathologique,* Paris, Presses Universitaires de France.

Christensen, W.D. 2012, 'Natural sources of normativity'. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43, 104-112.

Christensen, W.D., & Bickhard, M.H. 2002, 'The Process Dynamics of Normative Function', *The Monist* 85, 1, 3-28.

Craver, C.F. 2001, 'Role functions, mechanisms, and hierarchy', *Philosophy of Science* 68, 53–74.

Cummins, R. 1975, 'Functional analysis', *Journal of Philosophy* 72, 741–765.

Cummins, R. 2002, 'Neo-Teleology', en Ariew, A., Cummins, R. & Perlman, M. (Eds). *Functions*. Oxford, Oxford University Press, 157-172.

Davies, P. S. 2001, *Norms of Nature. Naturalism and the Nature of Functions,* Cambridge, MIT Press.

Davies, P. S. 2009, 'Conceptual Conservatism: The Case of Normative Functions', in Krohs, U. & Kroes (eds.) *Functions in Biological and Artificial Worlds.* Cambridge, MIT Press, 127-146.

Fraansen, M. 2009, 'The Inherent Normativity of Functions in Biology and Technology', in Krohs, U. & Kroes (eds.) *Functions in Biological and Artificial Worlds,* Cambridge, MIT Press, 103-126.

Godfrey-Smith, P. 1994, 'A modern history theory of functions', *Noûs* 28, 344–362.

Hume, D. 1978, *A treatise of human nature,* Oxford, Clarendon Press. Publicado originalmente en 1739.

Krohs, U. 2010, Dys-, 'Mal- et Non-: L'autre face de la fonctionnalité', en Gayon, J. & De Ricqlès. *Les fonctions: des organismes aux artefacts,* Presses Universitaires de France, 337-352.

McLaughlin, P. 2009, 'Functions and Norms', in Krohs, U. & Kroes (eds.) *Functions in Biological and Artificial Worlds,* Cambridge, MIT Press, 93-102.

Millikan, R.G. 1989, 'In defense of proper functions', *Philosophy of Science* 56, 288–302.

Mossio, M., Saborido, C. and Moreno, A. 2009, 'An Organizational Account for Biological Functions', *British Journal for the Philosophy of Science* 60 (4), 813-841.

Neander, K. 1995, 'Misrepresenting and malfunctioning', *Philosophical Studies* 79, 109-141.

Price, C. 1995, 'Functional Explanations and Natural Norms', *Ratio (New Series)* 7, 143-160.

Nordenfelt. L. 1987, *On the Nature of Health: An Action-Theoretic Approach,* Dordrecht, Reidel.

Reznek. L. 1987, *The Nature of Disease,* London, Routledge and Kegan Paul.

Saborido, C., Mossio, M & Moreno, A. 2010, 'La dimensión teleológica del concepto de función biológica desde la perspectiva organizacional', *Teorema* 29 (3), 31-56.

Saborido, C., Mossio, M & Moreno, A. 2011, 'Biological organization and cross-generation functions', *British Journal for the Philosophy of Science* 62, (3): 583-606.

Schlosser, G. 1998, 'Self-re-production and functionality: A systems-theoretical approach to teleological explanation', *Synthese* 116, 303–354.

Schramme, T. 2007, 'A Qualified Defence of a Naturalist Theory of Health', *Medicine, Health Care, and Philosophy* 10, 11–7.

Schwartz, P. 2007, 'Defining Dysfunction: Natural selection, design, and drawing a line', *Philosophy of Science* 74, 364-385.

Wakefield, J. C. 1992, 'The concept of mental disorder: on the boundary between biological facts and social values', *American Psychologist* 47, 373-388.

Wright, L. 1973, 'Functions', *Philosophical Review* 82, 139-168.

LA DISCUSIÓN DEL POSITIVISMO LÓGICO EN LA JAE: BLAS CABRERA Y JULIO PALACIOS

MARGARITA SANTANA DE LA CRUZ
Universidad de La Laguna
msantana@ull.es

ABSTRACT: The so-called "Silver Age" of Spanish science is inextricably linked to the JAE. The scientific production of the members of this institution, a privileged elite, came to Spain at the forefront of the science of time. This elite was actively involved in internationals discussions and forums, and the exchange of culture and science in other countries led to the opening, not only to science but to the currents of thought circulating in Europe and North America, including the philosophical. Thus, the objective of this paper is to show how openness and sharing in these forums not only resulted in the advancement of science but at the reception (introduction and circulation) of these currents of philosophical thought by scientists themselves. Epistemological and methodological concerns, motivated by interest in to do science with "modern methods" were part of the stage that was conceived and developed the effort to reach international standards in the production of modern science in Spain.
KEY WORDS: Science, JAE, Modernity, Positivism, Methodology, Empiricism.

La denominada “Edad de Plata” de la ciencia española está indisolublemente ligada a la JAE. La producción científica de los integrantes de esta institución, una élite privilegiada, acercó a España a la vanguardia de la ciencia del momento. Dicha élite participó activamente en los debates y foros internacionales, y el intercambio con la cultura y la ciencia de otros países propició la apertura, no sólo a la ciencia sino a las corrientes de pensamiento que circulaban por Europa y Norteamérica, incluidas las filosóficas. El objetivo del presente trabajo, por tanto, es mostrar cómo la apertura y el intercambio en esos foros no sólo se tradujeron en el avance de la ciencia sino en la recepción, introducción, circulación y, en ocasiones, discusión de dichas corrientes de pensamiento filosófico por parte de los mismos científicos. Las preocupaciones epistemológicas y metodológicas, motivadas por el interés en hacer una ciencia con “métodos modernos”, formaron parte del escenario en que se gestó y desarrolló el esfuerzo por alcanzar los estándares internacionales modernos en la producción de ciencia en España.

Desde casi mediados, pero sobre todo a finales, del siglo XIX se introducen en el país, después de una práctica hegemonía del krausismo como filosofía sintetizadora –aunque habrá posiciones intermedias o de transición como el krausopositivismo-, el evolucionismo y el positivismo, sobre todo el de corte comtiano, ya que se intentará aplicarlos como herramientas de análisis –y también como generadoras de posibles soluciones para ello- del acaecer sociopolítico de España. Esto es, se aplicarán sobre todo en el terreno del pensamiento sociológico y político[1]. El atraso del país respecto a Europa es evidente, su decadencia también. En la regeneración del mismo –tarea urgente a abordar- la ciencia y sus valores jugarán un papel esencial. La cultura española tenía que robustecerse con el contacto con el pensamiento científico, con las ciencias experimentales y sus desarrollos y resultados.

Desde el punto de vista filosófico la cuestión será conciliar razón y experiencia, ciencia positiva y especulación filosófica. La reflexión filosófica positiva obtiene su legitimidad a partir de los resultados de la experiencia científica. Se trata de encontrar un término medio entre el empirismo rígido y la pura especulación idealista –divorciada del desarrollo de la ciencia.

Cuando hablamos del positivismo y la JAE, sin embargo, no hablamos sólo de esto: hablamos de una metodología, de una epistemología, y de una forma de entender la ciencia, sus objetivos y su praxis. Así, en la producción científica de los miembros de la Junta nos encontramos, junto a los trabajos científicos específicos de la disciplina en la que investiga cada uno de ellos, una serie de obras que podríamos denominar de "filosofía de la ciencia" o de "metodología de la ciencia". España ha estado al margen de los desarrollos científicos de Europa, siendo una enorme desconocedora de las teorías, métodos y técnicas de la ciencia reciente, con lo cual es urgente la tarea de la modernización, que no sólo tiene que ver con la formación específica y técnica de un personal capaz de transformar la situación, sino con la comunicación de lo que es la ciencia y de cómo debe abordarse la investigación, con los supuestos metodológicos y epistemológicos que comporta. Frente a una concepción de la ciencia –ciencia importada- que subraya casi exclusivamente la pertinencia de la misma en términos de su aplicación, la necesidad de elaborar ciencia teórica; frente a la primacía de la especulación y los principios, la atención a los hechos, a la observación y a la experiencia.

Blas Cabrera y Julio Palacios, aunque también Rey Pastor, harán aportaciones sustantivas en esta dirección. El antecedente de las mismas es la obra de Ramón y Cajal, *Reglas y consejos sobre investigación científica. Los tónicos de la voluntad,* cuyo origen fue su discurso de ingreso en la Academia de Ciencias Exactas, Físicas y Naturales en 1897. En dicha obra el autor, además de otros

[1] Vid. Núñez Ruiz, D. 1975, *La mentalidad positiva en España: desarrollo y crisis,* Madrid, Túcar Ediciones.

temas como los deberes que el Estado tiene con la producción científica, desarrolla una exposición de lo que es y debe ser la investigación científica, una exposición de lo que constituye la metodología de la ciencia, unida a consideraciones fundamentales respecto al lenguaje y la comunicación de la ciencia. Dicha exposición se articula en torno a una serie de supuestos o prescripciones cuya observancia aparece como esencial, para el adecuado desarrollo de la misma, en el caso español.

En primer lugar, una crítica a lo que denomina "el culto excesivo" a la ciencia llamada práctica, subrayando la necesidad de la investigación teórica. En segundo lugar, la necesidad de la especialización frente al "enciclopedismo", encerrado en la máxima vulgar de "quien mucho abarca poco aprieta". En tercer lugar, y en relación con ello, la necesidad también de lecturas especiales o técnicas de autores originales. En cuarto lugar, la indispensabilidad y viabilidad de compatibilizar el ejercicio profesional y la labor investigadora. Y en quinto y último lugar, en estrecha relación con la metodología empleada, la redacción del lenguaje científico. La comunicación científica es condición sine qua non para el desarrollo de la ciencia, lo que incluye la especificidad de su lenguaje y de su estilo frente a los de otros modos de conocimiento.

Dice Cajal:

> "El estilo de nuestro trabajo será genuinamente didáctico; sobrio, sencillo, sin afectación, y sin acusar otras preocupaciones que el orden y la claridad.[…] El escritor científico aspirará constantemente a reflejar la realidad objetiva con la perfecta serenidad e ingenuidad de un espejo(…) y abandonando, en fin, la pretensión de estilista exquisito y el fatuo alarde de profundidad filosófica" (Ramón y Cajal 1897, 153).

Porque, en efecto, considera que "la flatulencia retórica" es "(la) manifestación del meridionalismo superficial y causa poderosa de nuestro atraso científico" (Ramón y Cajal 1897, 154). La publicación de la ciencia, además, debe llevarse a cabo en revistas extranjeras, garantizando así el intercambio genuino de ideas.

Este conjunto de supuestos o prescripciones, como decía, se articulan en torno a lo que es y debe ser la forma de proceder en ciencia si queremos alcanzar, pues tal es el objetivo, la verdad científica. Los pasos serían los siguientes:

> "1º. Observación de los hechos demostrados, a favor de métodos terminantes, claros y de gran precisión. 2º. Experimentación para crear condiciones nuevas en la manifestación de los fenómenos. 3º. Crítica y eliminación de una interpretación racional de los hechos, en cuya virtud éstos queden subordinados a una ley general y, si es posible, a una representación o esquema físico-químico. 4º. Comprobaciones de la hipótesis mediante nuevas observaciones o repetidos experimentos. 5º. De no concordar con la realidad, sustitución de la hipótesis por otra, que será a su vez sometida a riguroso análisis objetivo. 6º. Aplicaciones y

> ramificaciones de la hipótesis, ya convertida en verdad firme, a otras esferas del saber" (Ramón y Cajal 1897, 144).

El magisterio de Cajal no sólo es moral y de estímulo, es profesional e investigador, y en la línea abierta por él, en la que se insiste en que España sólo logrará salir de su atraso si sigue estas reglas y consejos, que son sobre todo metodológicas, se sitúa la obra de Blas Cabrera de 1936: *Evolución de los conceptos físicos y lenguaje;* y la de Julio Palacios de 1953: *El lenguaje de la física y su peculiar filosofía*. Enormes conocedores de la metodología científica, o del hacer la ciencia "con métodos modernos", también lo son de la reflexión filosófica sobre la misma.

Así, en el caso de Blas Cabrera nos encontramos, junto a un análisis del lenguaje científico, en concreto el relativo a la necesidad de su estandarización en nuestro idioma, cuestiones de metodología y de las implicaciones que éstas tienen para una teoría del conocimiento.

Al estilo de Cajal, pero teniendo en cuenta el tiempo que separa a ambas obras, Cabrera inicia su discurso aludiendo a la polémica de la ciencia española y a cómo fue cancelada por Cajal y por la JAE[2]. Su relato sobre el punto de inflexión que supone esta institución va acompañado por un recorrido por la historia de la física en España, con una universidad que ignoraba el método experimental. Como acabamos de señalar, dos serán, sin embargo, los temas principales abordados en esta obra: el relativo al lenguaje de la ciencia y el referente a la metodología científica y a las implicaciones que tiene para una teoría del conocimiento. Respecto al primero, Cabrera insiste en la necesidad de estandarizar el lenguaje científico y en los problemas que han surgido en nuestro idioma por la ausencia de dicha estandarización: careciendo de conceptos claros, el desconocimiento de la lengua en la que se hacía y divulgaba la ciencia supuso que el lenguaje español se viera invadido por el vocabulario de una técnica que es consecuencia del usufructo de sus ventajas materiales, pero no de la contribución a su progreso. En este sentido la idea nuevamente, como vimos en Cajal, es la crítica a la sobrevaloración de la ciencia práctica:

> "en la industria como en las demás actividades humanas lo esencial son las ideas, cuya realización se encomienda a los organismos sociales adecuados para darles efectividad" (Cabrera Felipe 1936, 21).

Nuestro idioma, en cualquier caso, ha sufrido un enorme daño por haber estado alejados del comercio intelectual en la época en que nacieron y se construyeron los capítulos fundamentales del saber. En la física, en concreto, "no sólo faltaron nombres adecuados para la designación de objetos

[2] A fines del siglo XIX tiene lugar lo que se denominó "la polémica de la ciencia española", una polémica que enfrentaba a los que consideraban que los españoles no habían hecho ningún tipo de aportación al escenario internacional de la ciencia, y a los que defendían , retrotrayéndose hasta la ciencia árabe, que sí lo habían hecho. En el caso de las matemáticas, por ejemplo, la polémica estuvo personificada en el primer caso por Echegaray, y en el segundo por Menéndez Pelayo. Vid. García Camarero, E. y E. 1970, *La polémica de la ciencia española,* Madrid, Alianza.

reales o conceptos que no cabían en los moldes hechos" (Cabrera Felipe 1936, 22). Respecto al segundo, el recorrido que realiza el autor por la historia de la física, desde Newton hasta Einstein y la física cuántica, le sirve para ilustrar, metodológicamente hablando, el modo de proceder en ciencia: la ciencia

> "(…) habrá de someterse a las sentencias inapelables de la experiencia. Una contradicción manifiesta de los fenómenos observados con las predicciones teóricas no podrá resolverse sino retocando aquellos postulados que erróneamente se creyeron reflejo exacto de la realidad*"* (Cabrera Felipe 1936, 25).

Esta metodología, a su vez, viene acompañada por una teoría del conocimiento concreta:

> "nuestro conocimiento no puede tener otras fuentes que la observación del mundo exterior mediante los órganos a ello destinados. Todas las magnitudes inobservables deben ser rechazadas; los únicos elementos con que podemos y debemos operar en nuestros cálculos son números que miden cualidades directamente observables" (Cabrera Felipe 1936, 54).

Y en relación con lo que supone el principio de indeterminación también afirmará:

> "¿Tiene o no sentido aceptar una realidad fuera de nuestro conocimiento? (…) En el nuevo punto de vista si se postula una realidad es con la condición de declararla inabordable" (Cabrera Felipe 1936, 57).

Cae fuera de las competencias de un físico, en cualquier caso, resolver el viejo problema filosófico de la realidad del mundo exterior.

En el caso de Julio Palacios, quizá más interesante por ser más explícito, el autor abordará también la temática del lenguaje en la ciencia pero en diálogo directo y discusión abierta con el positivismo lógico en su versión operacionalista. Planteando como objetivo principal de su trabajo determinar y clarificar el lenguaje de la física, en concreto las buenas definiciones en física, empezando por la propia definición de la disciplina, defenderá, frente al operacionalismo y el instrumentalismo de Bridgman y Mach, que el lenguaje de la física se corresponde con lo que denomina "el realismo ingenuo". Así,

> "(…) la ingenuidad puede llevarse hasta el extremo de creer que las masas, las fuerzas, las distintas formas de energía, existen antes e independientemente de que los físicos las hayan definido y medido*"* (Palacios Martínez 1953, 36).

Por esa razón "la física logra un lenguaje sencillo y suficientemente preciso a condición de que se atenga al realismo ingenuo" (Palacios Martínez 1953, 36)*,* que es la filosofía propia del físico.
La cuestión, a su juicio, es que, aunque los distintos científicos –físicos- pueden sostener opiniones filosóficas o metafísicas distintas al definirse –Planck y Enistein son realistas, Mach y Bridgman son operacionalistas-, en tanto que físicos son realistas ingenuos y no pueden no serlo. El peligro

del positivismo lógico es que se trata de un movimiento que, de triunfar, alteraría radicalmente el lenguaje de la física. Y aunque surgió con el propósito de evitar o prevenir la creación de teorías vanas, el seguimiento de sus directrices principales nos conduciría a intentar evitar continuamente el errar, cuando en realidad en ciencia no hay descubrimiento sin aventura.

Pero, ¿cuáles son los rasgos esenciales de este movimiento que tan peligroso o nocivo le parece al autor? Merece la pena recoger aquí, pese a su extensión, los rasgos que él mismo señala:

> "1. Las cantidades tales como la longitud, la masa, la temperatura y la carga eléctrica de un cuerpo determinado (…) han de definirse como el resultado objetivo de ciertas operaciones que, cuando menos en principio, han de poder ser realizadas en el laboratorio.
> 2. Las definiciones operacionales, lo mismo que todos los experimentos llevados a cabo en los laboratorios, conducen a un conjunto de datos instrumentales, tales como las lecturas por índices móviles sobre escalas (point readings). Estas lecturas (…) constituyen 'los hechos irrefutables de la naturaleza'. Aquellas cantidades para las cuales no pueda darse definición operacional, es decir, que no pueden reducirse a un sistema de lecturas, son 'inobservables', es decir, no pertenecen a la física.
> 3. Las leyes físicas son relaciones entre cantidades definidas operacionalmente, esto es, parecen quedar satisfechas cuando se realizan determinados experimentos.
> 4. El papel de la teoría se reduce a, mediante ciertas hipótesis, dar una descripción simple del mayor número posible de experimentos.
> 5. En lógica positivista no hay lugar para conceptos tales como 'verdad' y 'realidad'. Su propósito es lograr una descripción tan económica como sea posible de las percepciones sensoriales. Las mejores teorías son aquellas con las que se consigue mayor economía, por lo que, para los lógicos positivistas, economía es sinónimo de realidad" (Palacios Martínez 1953, 42).

De acuerdo con todo ello, las finalidades propias de la ciencia –y de la física-, como la explicación y la comprensión, dejarían de tener sentido. El único propósito legítimo sería el de la predicción, entendiéndola en relación, no con la observación directa de los fenómenos de la naturaleza, sino con la ejecución de determinados experimentos que deberían producir determinadas lecturas en los aparatos. Por otra parte, y en lo que se refiere al lenguaje de la física, la adopción del positivismo lógico así entendido nos llevaría a aceptar como válidas sólo las definiciones operacionales o experimentales, quedando excluidas las definiciones epistémicas. Sin embargo, un físico no puede admitir que cuestiones como la definición del calor o de la fuerza sean extracientíficas por ser entidades inobservables las que se definen. Así, el autor afirma:

> "La lógica positivista elimina del pensamiento los conceptos de verdadero y de real y se contenta con afirmar que las teorías son provisionalmente satisfactorias. Pues bien, yo no digo que la doctrina positivista sea verdadera ni falsa, pero, como físico, afirmo que no me satisface" (Palacios Martínez 1953, 45).

Metodología distinta para la ciencia desarrollada esos años en España, concepción del conocimiento y de la propia ciencia también distintas, todo ello será objeto de crítica una vez cancelado el proyecto de la JAE con la Guerra Civil y la posterior creación del CSIC. Pero, con todo, resulta innegable el papel jugado por aquélla como vía de transmisión no sólo de los desarrollos científicos que tenían lugar fuera de nuestras fronteras, sino también de la reflexión filosófica sobre los mismos.

REFERENCIAS

Cabrera Felipe, B. 1936, *Evolución de los conceptos físicos y lenguaje,* Madrid, C. Bermejo Impresor.

García Camarero, E. y E. 1970, *La polémica de la ciencia española,* Madrid, Alianza.

Gómez, A. y Canales, A. (eds.) 2009, *Ciencia y Fascismo,* Barcelona, Laertes.

Laporta, F. J. et alii, 1987 'Los orígenes culturales de la Junta para la Ampliación de Estudios'. *Arbor* nº 493, pp. 17-87, y nº 499-500, 9-137.

Núñez Ruiz, D. 1975, *La mentalidad positiva en España: desarrollo y crisis,* Madrid, Túcar Ediciones.

Palacios Martínez, J. 1953, *El lenguaje de la física y su peculiar filosofía,* Madrid, Real Academia Española.

Ramón y Cajal, S. 1897, *Reglas y consejos sobre investigación científica. Los tónicos de la voluntad,* Madrid, Espasa-Calpe, 2007.

Santana de la Cruz, M. 2009, 'Unidad de la patria, unidad de la ciencia: la retórica científica del régimen franquista', en Amparo Gómez y A. Canales (eds.), *Ciencia y Fascismo,* Barcelona, Laertes, 165-184.

VIOLATION OF *Measurement Independence* WITHOUT CONSPIRACY

I. SAN PEDRO
U. Complutense de Madrid

ABSTRACT: The aim of this paper is to show that *Measurement Independence*-type conditions are not sensible assumptions to require in the context of common cause explanations of the EPR correlations, and must thus be disposed of. I will argue, on the one hand, that *Measurement Independence* does not adequately express the requirement that EPR experimenters have free will, as it is usually claimed. On the other hand, I will suggest that actual measurement operations play a crucial causal role in any common cause account of EPR one may think of. This that such causal explanations of EPR of course means violate *Measurement Independence*.

Key words: *Measurement Independence*, Free will, Common Causes, EPR.

1 Introduction

Ever since Bas van Fraassen's influential "Charybdis of Realism ..." (van Fraassen, 1982) it has become a widespread opinion among philosophers that common cause accounts of EPR violations are to be ruled out. This is because the idea of *common cause* Reichenbach (1956) can be identified, as van Fraassen showed, with that of *hidden variable* entering Bell's theorem (Bell, 1964). Postulating the existence of (screening-off) common causes for the EPR correlations thus leads to the Bell inequalities, which are known to be empirically violated.

Derivations of the Bell inequalities in this context usually presuppose a common cause —as a hidden variable— on to which several constraints are set. Constraints on the postulated common causes are intended to reflect standard requirements on a generic physical system, including temporal order of causal relations or locality considerations. The strength of such arguments of course relies on the plausibility of the conditions imposed on the common causes. One such condition corresponds to the requirement that EPR experimenters do take free independent decisions at the moment of setting up the EPR apparatus for measurement. This is usually expressed by means of the so-called *No-conspiracy* condition —here I shall refer to this condition, more neutrally, as *Measurement Independence*.

The aim of this paper is to show that *Measurement Independence*-type conditions are not sensible assumptions to impose on the idea of common cause when aiming at causal explanations of

the EPR correlations. I will suggest therefore that we must dispose of such conditions in the EPR context. I will argue, on the one hand, that *Measurement Independence* does not adequately express the requirement that EPR experimenters have free will, as it is usually claimed. On the other hand, I will suggest that actual measurement operations play a crucial causal role in any common cause account of EPR we may think of. This results as well in the violation of *Measurement Independence*.

2 *Measurement Independence* and Free Will

Measurement Independence —i.e. *No-conspiracy*— is the requirement that common causes C postulated to explain the EPR correlations be probabilistically independent from the corresponding measurement operations m_i performed on either wing of the experiment, i.e.

$$p(m_i|C) = p(m_i). \tag{1}$$

(We write m_i for a generic measurement operation in an EPR experiment, with $i = L, R$ indicating that measurement is performed on the left and right wings of the experiment respectively.)

That equation (1) must hold in any common cause explanation of the EPR correlations is usually justified from the fact that EPR experimenters act freely to choose which specific measurements to perform each time. In what follows I shall argue against this very view, i.e. against the justification of equation (1) as a requirement related to the preservation of free will. In particular, what I shall challenge is what I take to be the general agreement that *Measurement Independence* is necessary for *free will*, i.e.

$$\text{Free will} \rightarrow p(m_i|C) = p(m_i). \tag{2}$$

Defenders of *Measurement Independence* as a sensible constraint to demand in the context of common cause explanations of EPR correlations may quickly point out that this feels really like beating up a straw man. For, they would claim, it is not free will actually what is behind the justification of equation (1), but just the intuition that the world is not such that it conspires to pre-set the measurement choices in the EPR experiment, regardless of how are those decided.[1] We could call this the "no-conspiracy justification" approach. The point being that free will is not necessarily involved in the requirement that the decisions about what measurement is to be performed in an EPR experiment need to be independent of the hidden (common cause) variable postulated.

I see two ways to defuse such claims. First, as I have already argued elsewhere (San Pe-

[1]This is, by the way, where the origin of the terminology *No-conspiracy* can be traced back to. However, since I will defend that violations of equation (1) do not constitute conspiracies of any kind, I shall use the less prejudiced more neutral expression *Measurement Independence*.

dro, ming), there are good reasons to think that physicist and philosophers alike seem to endorse, and implicitly assume, that free will indeed plays a role in the justification of *Measurement Independence*-type conditions. Prominent examples include John Bell himself (Bell, 1981), philosophers such as Huw Price (Price, 1994), or more recent discussions related to the so-called "Free Will Theorem" (Conway & Kochen, 2006). Still, this might be not enough for defenders of the "no-conspiracy justification" approach. In that case, I should point out as well that the arguments I provide in what follows aimed at detaching the idea of free will from that of *Measurement Independence* can be straightforwardly generalised to show that the appeal to (the lack of) world conspiracies does not provide either a sound justification of *Measurement Independence*.

3 *Measurement Independence* is not Necessary for Free Will

I shall pay attention here to three specific presuppositions behind *Measurement Independence*, all related to causal considerations the idea of free will seems to involve. Note that these might not be the only presuppositions at stake. I do take them however to be at the intuitive core of *Measurement Independence* as an assumption about free will.

First, the requirement of *Measurement Independence* as a condition about free will is based on the idea that freedom of will entails the lack of certain causal relations. More precisely, the probabilistic independence expressed by *Measurement Independence* is taken to represent causal independence. However, for a probabilistic expression such as *Measurement Independence* to stand as a causal statement at all, we need to assume at least that there is indeed a (faithful) correspondence between causal relations and probabilistic relations. Let us call this the *Cause-statistics Link* assumption.

This is a presupposition that can be easily challenged. In fact, there are many counterexamples that show that probabilistic dependence/independence is not necessary for causation, and certainly not sufficient either. Only in some cases and under certain "good" conditions can the *Cause-statistics Link* assumption be considered adequate. So we could conclude that *Measurement Independence* is not necessary for free will just by rejecting the idea that causal relations are adequately expressed in terms of probabilistic relations. This move would have however undesired consequences. For instance, if our analysis is motivated to some extent by the desire of explaining the EPR correlations causally —making use of the Principle of the Common Cause for example— we could not afford rejecting the necessity claim (2) on the grounds above. For if common cause explanations are to make sense at all then the *Cause-statistics Link* presupposition needs to be in place. Thus, while rejecting the *Cause-statistics Link* would undermine the claim that *Measurement Independence* is necessary for free will, it would also eliminate any possibility of explaining the EPR correlations

in terms of common causes.

A further issue as regards *Measurement Independence* is that it is the result of demanding, not only the lack of *some* causal influence between the postulated common cause and the events representing the setting of the experiments (and therefore between the common cause and the experimenters' decisions), but the lack of *all* causal influences between these. This I will refer to as the *No-cause* presupposition.

We may soon realise that *No-cause* may turn out to be too strong a requirement. For demanding no causal influence *at all* seems to suggest either a deterministic causal view as regards the (hidden) common cause events, or at least an idea of cause that exhausts all possible causal factors of a given effect. But it certainly is a possibility, specially in indeterministic contexts, to conceive the postulated common cause C to be just a *partial* cause of the measurement setting events m_i. Obviously *Measurement Independence* would not hold in this case. Neither would that constitute a violation of free will. To the contrary, it seems very sensible to think that free will would still be preserved even in the case our range of choices, or acts, had been somehow limited. So, in this view, free will is again completely compatible with the violation of *Measurement Independence*.[2]

Finally, the requirement of *Measurement Independence* in the EPR picture also presupposes a certain time ordering of the events involved. Common causes, in particular, are assumed to take place *before* measurement operations do. Let us call that the *Time Order* presupposition.

Note first that the precise time arrangement that the *Time Order* presupposition demands makes sense only in the context of a further causal assumption, namely that *cause events are in the past of their effects*. Thus requiring *Measurement Independence* involves a combination of an assumption about the *temporal order* of events as well as an assumption about the correct *causal order* to be taken. Both these two presuppositions can be contested independently. This leads to at least three different (causal) pictures, depending on which of the two assumptions above is dropped.

One may want to keep in the first place the temporal arrangement of events as it is usual. That is, one may want to stick to the idea that the postulated common cause is to take place before *both* the measurement operation events (in both wings) and the outcome events. The intuition is then that for violations of *Measurement Independence* not to entail a lack of free will we need to endorse the possibility of backwards in time causation. In particular, actual measurement operations m_i can

[2]This latter option is not available in deterministic contexts, though. For the presence of *any* cause entails there (with probability one) the occurrence of the corresponding effect. Defenders of the idea of free will in deterministic contexts should radically change their strategy and endorse for instance *compatibilism* and reduce free will to a psychological subjective feature of ours. There is no need to revise the *compatibilism-incompatibilism* debate here. It suffices to stress the fact that there are several options available to be able to say that *No-cause* is too strong an assumption, also in deterministic contexts.

Figure 1: *Causal picture 1*. Time order of events is preserved but causes propagate backwards in time.

Figure 2: *Causal picture 2*. Time order of events is *not* preserved and causes propagate, as usual, forward in time.

be thought to be (future) causes of the postulated common cause C, which would cause in turn the measured outcomes O_i (see Figure 1). If the common cause C is located in a sufficiently distant past, this picture turns out to be completely local, hence avoiding the usual conflicts with special relativity. Needless to say that the appeal to backwards causation is taken by many as a highly counterintuitive option. But it is still an option, as some have convincingly argued (Price, 1994).

A second causal picture results from rejecting the usual temporal arrangement of events while keeping the assumption that causes propagate forward in time to cause their effects (see Figure 2). In this case, the postulated common cause C can be thought to take place sometime in between the actual measurement operations m_i and the occurrence of the observed outcomes O_i. Measurement is taken to be in this view an explicit causal factor (of both C and the outcomes), hence implying that *Measurement Independence* is violated. As pointed out, this causal picture retains the most accepted intuition that causes propagate forward in time. This comes at a price nevertheless. Namely, a common cause model built along these lines seems to forcefully involve (explicit) non-local influences.[3]

Finally, a third causal picture would result form the rejection of both the usually presupposed time order of the EPR events as well as the intuition that causation propagates forward in time. If

[3]See (San Pedro, 012a) for a detailed discussion of such a model.

Figure 3: *Causal picture 3*. Time order of events is *not* preserved and causes propagate backwards in time.

backwards causation is again brought into the picture, the time order of events can be easily rearranged such that the postulated common cause C is in the future of the EPR outcome events. The common cause may even be thought to be situated far enough in the future so as to guarantee that causal interactions be completely local (see Figure 3).

In sum, the specific time order of events considered as well as our assumptions about the direction of causation are crucial in assessing violations of *Measurement Independence*, which need not involve a lack of freedom of will on the part of the EPR experimenters, nor a world conspiracy.

4 Causal Relevance of Measurement in EPR

Despite their obvious differences the three causal pictures above all share an important common feature. Namely they all conceive and postulate common causes as tightly related to measurement operations. This, as we saw, has different consequences —either in the form of backwards in time causation or of non-local interactions— which constitute a clear challenge to our standard intuitions. We should perhaps ask then whether the causal pictures above can really be useful at all as explanations of EPR correlations.

In what follows I shall give a positive answer to this question. I shall suggest that measurement operations play a specific causal role in the EPR picture, which is very much related to the role common causes should also play. I shall argue, in particular, that if a measurement operation in one wing of the experiment is causally relevant to the outcome in that same wing, then the very same operation of measurement will be causally relevant to the outcome in the other (distant) wing. More precisely, if m_L represents a measurement operation performed in the left wing of the experiment, and O_L and O_R the corresponding left wing and right wing outcomes, I claim the following holds:

Claim 1. *If measurement operation in the left wing m_L is* causally relevant *to the corresponding wing's outcome O_L, then it is also causally relevant to the distant (right) wing's outcome O_R.*

To show that the above holds, let me point out first that it does not seem problematic, at least form an intuitive point of view, to say that, taking each wing of the experiment individually, measurement operations are causally relevant to the whatever outcome we happen to measure in it. In other words, I take it as uncontroversial to say for instance that measurement operations on the left wing of an EPR experiment m_L are *causally relevant* to the outcomes O_L we happen to register in that wing of the experiment. This can be expressed probabilistically as follows:

$$p(O_L|m_L) \neq p(O_L|\neg m_L). \tag{3}$$

Now we know that there are specific situations for which the outcomes of an EPR experiment are *perfectly correlated*. Say for instance this is the case for specific instances of O_L and O_R, i.e. $p(O_L|O_R) = 1$. This can also be written

$$p(O_L \wedge O_R) = p(O_L) = p(O_R). \tag{4}$$

Using equations (3) and (4), it is easy to show that[4]

$$p(O_R|m_L) \neq p(O_R|\neg m_L). \tag{5}$$

Equation (5) we just derived has a straight causal reading, very much along the lines of expression (3). It says, in particular, that the event m_L representing a measurement operation in the left wing of the EPR experiment is *causally relevant*, at least for the case of perfectly correlated spins, for the outcome O_R measured on the right wing.[5] (A completely symmetrical argument would entail that measurement operations in the right wing of the experiment m_R are also *causally relevant* for the corresponding left wing outcomes O_L.)

In sum, measurement operations m_L and m_R can be shown to be *both* causally relevant to *both* the outcomes O_L and O_R in an EPR experiment when these are perfectly correlated. This has a number of consequences. Most notably, the fact that this is the case for perfect correlated spin measurements is enough to conclude that we cannot require *Measurement Independence* to hold in the EPR context at all. This is because saying that both m_L and m_R are jointly causally relevant for both O_L and O_R is nothing more than saying that the measurement operations events are *common* causal factors of the correlated outcomes. It seems thus clear that both m_L and m_R will be tightly related to any common

[4]The proof is rather straightforward using basic probability calculus.

[5]The causal interpretation of equations (3) and (5) of course relies on what I called the *Cause-statistics Link* assumption.

cause we might think of to explain the corresponding correlation. We should thus expect to find probabilistic dependences between measurement operations and the corresponding common causes. This means in turn that *Measurement Independence* is violated, i.e.

$$p(m_i|C) \neq p(m_i), \quad \text{with} \quad i = L, R. \tag{6}$$

This applies, as we saw, to perfectly correlated spins, but it can be easily generalised for non-perfect correlations. In particular, if measurement operations are casually relevant in the case of perfect EPR correlations, why should we think this is not going to be the case for non-perfect correlated spins? Both perfect and non-perfect correlations in EPR are, after all, observed as involving essentially the same kind of events. Thus, it seems most natural to think of measurement operations as causally relevant to the outcomes in *both* wings of the EPR experiment *also* in the case these are not perfectly correlated, as Claim 1 expressed.

5 Conclusions

I have shown that the commonplace claim by which *Measurement Independence* is taken to be necessary for the idea of free will in the context of causal explanations of EPR correlations is, although apparently correct according to certain intuitions, ultimately mistaken.

At least three underlying assumptions can be identified, all of which are revisable and can each be challenged on its own grounds. As a result, the notion of free will is shown to be compatible with the violation of *Measurement Independence* in different fashions, depending on which of the assumptions is rejected, and with diverse implications in each case. Most interesting are causal pictures featuring explicit causal relevance of measurement operations, since they may provide the grounds for actual common cause models of the EPR correlations.

Agradecimientos

Financial support from the Spanish Ministry of Science and Innovation, research project FFI2011-29834-C03-01, is gratefully acknowledged.

REFERENCES

Bell, J. S. (1964). On the einstein-podolsky-rosen paradox. *Physics*, *1*, 195–200. Reprinted in Bell (1987), pp. 14–21.

Bell, J. S. (1981). Bertlmann's socks and the nature of reality. *Journal de Physique*, *Colloque C2, suppl. au numero 3, Tome 42*, C2 41–61. Reprinted in Bell (1987), pp. 139–58.

Bell, J. S. (1987). *Speakable and Unspeakable in Quantum Mechanics*. Cambridge University Press.

Conway, J. & Kochen, S. (2006). The free will theorem. *Foundations of Physics*, *36*, 1441–1473.

Cushing, J. & McMullin, E. (Eds.). (1989). *Philosophical Consequences of Quantum Theory*. Notre Dame: University of Notre Dame Press.

Price, H. (1994). A neglected route to realism about quantum mechanics. *Mind*, *103*, 303–336.

Reichenbach, H. (1956). *The Direction of Time*. 1956. Unabridged Dover (1999) republication of the original University of California Press (1956) publication.

San Pedro, I. (2012a). Causation, measurement relevance and no-conspiracy in epr. *European Journal for Philosophy of Science*, *2*, 137–156.

San Pedro, I. (Berlin: Edition Open Access. Forthcoming). On free will and no-conspiracy. In Wüthrich, A. & Sauer, T. (Eds.), *Current Interpretational Problems of Quantum Theory: Proceedings of an International Symposium, Bern, June 23–24, 2011*. Max Planck Research Library for the History and Development of Knowledge., Berlin: Edition Open Access. Forthcoming.

van Fraassen, B. C. (1982). The charybdis of realism: Epistemological implications of bell's inequality. *Synthese*, *52*, 25–38. Reprinted with corrections in (Cushing & McMullin, 1989, pp. 97–113).

AN OUTLINE OF WEYLEAN SKEPTICISM*

IULIAN D. TOADER
Center for Logic, History and Philosophy of Science
Department of Philosophy, University of Bucharest.
itoad71@gmail.com

Abstract: This paper introduces Weylean skepticism, the view that objectivity and intelligibility are opposite ideals of science, explains the motivation for it and the argument that justifies it. The paper indicates also a couple of ways in which this skepticism could be resisted.

Keywords: objectivity, understanding, skepticism, phenomenology, axiomatics

Introduction

The goal of this short paper is to introduce you to Weylean skepticism. This is the view that modern science, to the extent that it strives to attain objectivity, i.e., to provide knowledge about the mind-independent reality, may do so only at the expense of intelligibility; and, vice versa, to the extent that it aims at intelligibility, may do so only by sacrificing objectivity. It was maintained, I believe, by the German mathematician and theoretical physicist Hermann Weyl, who thought that objectivity and intelligibility are fundamentally antagonistic epistemic ideals, which pull science in opposite methodological directions -- one driven by Husserl's pure phenomenology, the other by Hilbert's axiomatic formalism.

Weylean skepticism is important for a couple of reasons. First, although it underlies the most characteristic aspect of Weyl's philosophy of science, it seems not to have been recognized by contemporary scholars working on Weyl. Secondly, a careful analysis of this kind of skepticism should help us see what, if anything, would prove capable of dissolving the tension Weyl saw between objectivity and intelligibility, thereby giving support to the claim that scientific theories *can* help us to understand natural phenomena, as well as to justify their mind-independent reality.

* This paper stems from my doctoral dissertation (Toader 2011) written under the direction of Mic Detlefsen and Don Howard, whom I want to thank here again for great guidance and wonderful support. Financial support for writing the paper was provided through the 7th European Community Framework Programme (Marie Curie CIG No. 293899) and by the University of Bucharest. My attendance to the 7th Conference of the Spanish Society for Logic, Methodology and Philosophy of Science, at the University of Santiago de Compostela, where this paper was presented to a very nice and helpful audience, was made possible through a travel grant from the Romanian Cultural Institut in Madrid. Thanks to Ioana Alupoae and Concha Martínez Vidal for great organizational support.

Below, I indicate just two potential criticisms of Weylean skepticism: one via a pure phenomenological approach to objectivity, the other via a formal axiomatic approach to understanding. But these criticisms are not fully pursued here.[1]

1. Weyl on intuitionism, formalism, and phenomenology

Weylean skepticism can be best introduced, I believe, by starting from an argument that Weyl presented in 1927 at the University of Hamburg: "If Hilbert's view prevails over intuitionism, which in all appearance is the case, *then I see in this a decisive defeat of the philosophical position of pure phenomenology*, which thus proves to be insufficient for the understanding of creative science even in the area of cognition that is most primal and most readily open to evidence – mathematics." (Weyl 1928, 149) A careful analysis of this argument has to clarify his reasons for thinking that, in the late 1920s, formalism had prevailed over intuitionism, as well as his reasons for maintaining that this victory entails a defeat of pure phenomenology. This will allow us to see why he came to the view that objectivity and intelligibility are opposite epistemic ideals of science.

At a first glance, what seems to have provided substantial reasons for Weyl's thinking that Hilbert won the dispute with intuitionism is the belief, shared by many in the late 1920s, that a proof of the consistency of arithmetic was, or could be, found. In a 1929 paper, Weyl wrote that this is "a concrete mathematical problem which is not trivial, but at the same time is solvable," and he added that, in the case of classical analysis, the situation "remains serious, but not entirely hopeless." (Weyl 1929, 170) One might, accordingly, criticize Weyl's argument by pointing out, with hindsight, that the claim that Hilbert's formalism prevailed over intuitionism is simply false, for Goedel's second incompleteness theorem shows precisely that even weak fragments of arithmetic cannot prove their own consistency. Against this criticism stand, of course, the various attempts to show that (at least a revised form of) Hilbert's foundational program can overcome Goedel's challenge. As is well known, one such attempt proposed an extension of the type of reasoning that may be used in a consistency proof, beyond the limits envisaged by Hilbert (Gentzen 1936). Another attempt argues, on the basis of certain epistemic criteria, for a restriction of the mathematics that needs to be secured via a consistency proof (Detlefsen 1986). Yet another attempt is represented by the so-called "reverse mathematics" program (Simpson 1999).

But Weyl's reasons for his reconciliation with Hilbert are different than the ones motivating such attempts. Indeed, as he came to suspect that formalized mathematics is a mere game with symbols on paper and, thus, incapable of bringing about understanding, Weyl denied that a

[1] The interested reader is directed to the extended presentation and discussion of Weylean skepticism in Toader 2011.

consistency proof would be enough to vindicate formalism. Furthermore, in the face of Goedel's challenge, he argued that a consistency proof is not required and that the formalist commitment to finding one is unjustified. His reconciliation with Hilbert can be best explained by paying attention to Weyl's view about the role of mathematics in physics. It is this view, as I will briefly point out below, that led him to believe that objectivity and intelligibility are fundamentally antagonistic epistemic ideals of science.

2. Weylean Skepticism

The argument for Weylean skepticism, as presented in (Toader 2011), starts with an analysis of Weyl's view about the conditions for belief and understanding in mathematics. More particularly, it begins by clarifying his criticism of an idea defended by Dedekind, according to whom, in science, one should not believe a provable proposition without proof, even if that proposition is intuitively evident. He believed that although evidence may be the proper basis for believing an unprovable proposition, it is never the proper basis for believing a *provable* proposition. Weyl criticized this view by maintaining that evidence is the most proper basis for belief and by arguing that if a provable proposition is evident, then any attempt to prove it is useless. To clarify Weyl's criticism, one should note that Weyl adopted Husserl's conception that evidence is an experience (*Erlebnis*) of truth. An experience of truth, unlike a feeling of inescapable certainty that a proposition is true, is thought to be an epistemic achievement of the highest type, characterized by the complete satisfaction of the meaning intentions expressed in a proposition through the presentation in intuition of the intended objects. I think it is this phenomenological conception of evidence that allowed Weyl to allege that if a provable proposition is evident, then to properly believe that proposition one is not required to prove it. For, on this conception, if one has an experience of truth, then one has an unexcelled basis for belief.

But do I also think that, on Weyl's view, an experience of truth is also the proper basis for understanding why a theorem is true? Despite his endorsement of the view that mathematical objects are creations of the mind, he rejected the idea that evidence, conceived of as a feeling caused by one's creatorly experiences, is sufficient for understanding why a theorem is true. He seems to have come to believe, on the one hand, that a feeling of evidence cannot show *why* a theorem is true, since, as Husserl pointed out, it cannot even indicate *that* the theorem is true. On the other hand, Weyl seems to have believed that an experience of truth is indispensable for understanding, provided that such an experience is admitted as a form of proof. An experience of truth may, however, be only admitted as a proof that follows the method of what Weyl called

"immanent" axiomatics (as opposed to what he called "transcendent" axiomatics). In immanent axiomatics, mathematical reasoning is wholly contentual and proceeds from axioms formulated in terms of general concepts obtained by abstraction from what is immediately presented in intuition. According to Weyl, understanding requires an immanent axiomatic proof, i.e., a proof that provides a construction of the objects referred to in a theorem, and thereby completely satisfies the meaning intentions expressed in the theorem.

Weyl's view on mathematical understanding raises an important question about the scientific relevance of large parts of modern mathematics. He maintained, for instance, that proofs that follow the method of transcendent axiomatics -- proofs that include transfinite components, such as unbounded existential quantification over infinite domains -- fail to bring about understanding: not only can they not show why a theorem is true, they cannot even indicate that it is true. In transcendent axiomatics, concepts are obtained not by abstraction from what is presented in intuition, but freely created by the mind, and introduced as symbols through the positing of the axioms under certain constraints, like consistency. The use of such concepts, Weyl contended, entails that a transcendent axiomatic proof cannot be seen as an experience of truth, since part of its reasoning -- the part involving transfinite components -- is non-contentual or purely symbolic. Weylean skepticism is motivated by the belief that the use of such reasoning bars understanding, although, as I point out below, it is necessary for objectivity.

To see the latter, one should analyze Weyl's critical reflections on traditional empiricist approaches to modern science, in particular on Hobbes' constructivism. Weyl seems to have thought that any such approach is marred by reliance on empiricist abstraction, for this method does not allow the introduction of hypothetical elements, i.e., real but in principle unobservable entities. This is why, on his view, traditional empiricism collapses into phenomenalism, of the sort defended in Leibniz' monadology, which, according to Weyl, may account at most for intersubjectivity. One should also realize that Weyl was very enthusiastic about Fichte's constructivism, as developed in the *Wissenschaftslehre*. This view, however, as Weyl pointed out, fails to provide a correct account of the connection between freely created concepts and our experience, for it claims that what can be known through perceptual experience coincides with what is freely created by the mind and, thus, known independently of perceptual experience. This entails that the *Wissenschaftslehre* actually embraces monadological phenomenalism and, thus, according to Weyl, may offer only an inadequate account of objectivity. Finally, one should note that he was also drawn to Husserl's phenomenology, especially to his criticism of traditional empiricist abstraction, according to which the latter fails to distinguish between an intending act of the mind and an intended object. Husserl

argued that general concepts are introduced by abstraction from the relations between intending acts, rather than from the properties of intended objects. This method of concept formation -- the so-called "ideational" abstraction -- was used by Weyl, early in his career, for a development of Einstein's general relativity via a purely infinitesimal geometry. This development requires the elimination of concepts abstracted from the relations between intending acts that cannot be completely satisfied through the intuitive presentation of the intended objects (e.g., the concept of Riemannian congruence), justified by the fact that no scientific proposition formulated in terms of such concepts could allow an experience of truth. According to the young Weyl, a proposition is objective only if one knows that it remains true when relativized to different coordinate systems. For, if one takes a coordinate system as representing the perspective of transcendental subjectivity, as he did, then a proposition is objective only if one knows that it is true for any transcendental subjectivity. But since, according to Husserl, experiencing the truth of a proposition is enough for knowing that the proposition is true for any transcendental subjectivity, Weyl's view seems to have been that a proposition is objective only if one can experience its truth. However, he later rejected this view, coming to believe that the experience of truth cannot be a necessary condition for objectivity, since it entails the elimination of hypothetical elements. According to the mature Weyl, the assumption of real but in principle unobservable entities is indispensable for objectivity, as the only means for overcoming phenomenalism. Thus, Husserl's phenomenology also collapses into phenomenalism and may, therefore, just like Hobbes' and Fichte's constructivisms, support at most intersubjectivity.

Objectivity, Weyl came to believe, requires that scientific concepts (like force, energy, electromagnetic field, etc.) be freely created and introduced as mere symbols through the positing of fundamental theoretical principles. This is, I submit, the main belief that supports Weyl's Hamburg argument quoted above. As already noted, Weyl made a strong claim that purely symbolic reasoning, as in Hilbert's transcendent axiomatics, fails to bring about understanding. Nevertheless, Weyl also took the view that purely symbolic reasoning should not, although it might, be considered a mere game with symbols. For he maintained that mathematical concepts (including transfinite components) partake in the theoretical construction of the mind-independent world, and they do so in the same way as the scientific concepts (including hypothetical elements) do. In other words, Weyl's view seems to have been that, although it cannot have any *explanatory* role, the method of transcendent axiomatics has an *objectifying* role: it is indispensable for scientific objectivity, i.e., for the formation of theories that may provide knowledge about the mind-independent world. However, a scientific theory, according to him, may provide objective knowledge only if it is categorical, i.e.,

only if the connection between a theory and its domain of application is univocal up to isomorphism, where the criterion for univocality is expressed by the notion of concordance -- the idea that all methods for determining the value of a physical quantity must lead to the same result (within the limits of experimental error). This view on objectivity entails a fundamental limitation of scientific knowledge, in the sense that only the relations between mind-independent objects, and not these objects themselves, may be epistemically accessible through natural science. This limitation indicates again the tension that Weyl saw between scientific objectivity and intelligibility

Now, one way of criticizing Weylean skepticism is to resist Weyl's conclusion that a phenomenological approach to scientific objectivity is doomed to failure. Indeed, this has been attempted in recent literature (e.g., Ryckman 2005). Instead of sketching this line of argument, I should like to further emphasize that Weyl denied any connection between the free creation of concepts and scientific understanding. Thus, another possible way to criticize Weylean skepticism is to show that the tension that he saw between objectivity and intelligibility can be dissolved by arguing that partly non-contentual or purely symbolic reasoning, i.e., reasoning through scientific idealization, can render natural phenomena understandable. One condition that I think is required for understanding is that idealization be controlled via a proof of concordance, i.e., that one prove by contentual reasoning that the outcome of reasoning through idealization does not deviate from actual observations more than by what may be due to experimental error. But this claim faces at least two criticisms. First, one might deny that a concordance proof has any serious epistemic significance. In the face of the provability limitations suggested by Goedel's second incompleteness theorem, Weyl actually contended that the formalist commitment to proving logical consistency is unjustified, and suggested that the mathematical logician should merely attempt to restore consistency if a contradiction comes up, just like the physicist, who attempts to restore concordance only when this is found to be lacking. This contention indicates, however, that Weyl failed to see the epistemic significance of proving consistency and concordance. Without such proofs, the reliability of purely symbolic reasoning, as well as its ability to bring about understanding, would be lost. Secondly, one might claim that even if a concordance proof could be taken to control idealization, this would be insufficient for understanding. But I would point out that a further requirement for understanding is that reasoning through idealization be simpler, i.e., more thought-economical, than reasoning without idealization, and argue that if this further requirement is satisfied, then one is justified to deploy idealization in bringing about understanding if one can adequately justify as well the belief that science strives to bring about understanding in the most efficient way.

Conclusion

In this paper, I have briefly introduced Weylean skepticism, i.e., the view according to which objectivity and intelligibility are fundamentally antagonistic epistemic ideals of modern science. I explained what I take to be Weyl's own motivation for such a view, and what I think is the argument that justified it. Finally, I also indicated two ways in which this view could be resisted. Further work is, however, necessary for refuting this kind of skepticism, and so I hope that other philosophers of science will join in this work.

REFERENCES

Detlefsen, M. 1986, *Hilbert's Program: An Essay on Mathematical Instrumentalism*, Dordrecht, Kluwer.

Gentzen, G. 1936, 'Die Widerspruchsfreiheit der reinen Zahlentheorie' in *Mathematische Annalen* 112, 493-565. Eng. tr. in *The Collected Papers of Gerhard Gentzen*, Amsterdam, North-Holland, 1969, 132-213.

Ryckman, Th. 2005, *The Reign of Relativity: Philosophy in Physics 1915-1925*, Oxford Studies in Philosophy of Science, Oxford University Press.

Simpson, Stephen G. 1999, *Subsystems of Second Order Arithmetic*, Perspectives in Mathematical Logic, Springer, Berlin.

Toader, I. D. 2011, *Objectivity Sans Intelligibility. Hermann Weyl's Symbolic Constructivism*, ProQuest Information and Learning, Ann Arbor, Michigan.

Weyl, H. 1928, 'Diskussionsbemerkungen zu dem zweiten Hilbertschen Vortrag ueber die Grundlagen der Mathematik' in *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universitaet* 6, 86-88, reprinted in *Gesammelte Abhandlungen* III, 147-149. Eng. tr. in J. van Heijenoort (ed.) *From Frege to Goedel*, 482-484.

Weyl, H. 1929, 'Consistency in Mathematics' in *The Rice Institute Pamphlet* 16, 245-265, reprinted in *Gesammelte Abhandlungen* III, 150-170.

MECANISMOS, SISTEMAS DINÁMICOS Y EL SANTO GRIAL DE LA EXPLICACIÓN CIENTÍFICO-COGNITIVA

NICOLÁS VENTURELLI
Universidad Nacional de Córdoba / CONICET

ABSTRACT*:* I develop an advisory assessment of the recent debate concerning the impact of the dynamical approach on philosophical accounts of explanation for the cognitive sciences. The remarks here advanced are specially aimed at evaluating the role and relevance of the debate facing the growing attention that dynamicist research is harvesting. Towards this end, I argue for a greater appreciation of the relation between explanation and other cognitive goals, and an increased sensitivity towards the associated contextual factors.
KEY WORDS: Explanation, dynamical approach, scientific models, mechanisms, context of discovery.

Introducción

Durante los últimos años la cuestión de la explicación científica ha cobrado una renovada importancia en la filosofía de las ciencias cognitivas atendiendo especialmente a la mayor visibilidad del llamado enfoque dinamicista (ED) en las ciencias cognitivas (van Gelder & Port 1995, Beer 2000). En particular, la utilización extensiva, característica del enfoque, de la teoría de sistemas dinámicos (por ejemplo, Strogatz 1994) en la elaboración de los modelos parece poner en cuestión algunas asunciones relevantes respecto del tipo de explicación, y por ende el tipo de teorías, que deben ser procuradas en este campo de la investigación científica. A raíz de consideraciones relativas al aparato matemático adoptado para la construcción de los modelos, parece perfilarse un tipo de explicación alejado del que subyace a modelos más establecidos, expresados en términos de representaciones mentales o internas y procesamiento de la información.

En este trabajo siento algunas bases para un pronóstico en buena medida negativo sobre el desarrollo del debate en torno de cómo el ED viene a impactar sobre nuestra concepción de la explicación científico-cognitiva. Intento a la vez divisar algunas directivas para poner de relieve aquellos de sus aspectos que tengan injerencia sobre el diseño y la conducción de la investigación en este complejo campo científico. En este sentido, el que pretendo hacer es un aporte de tipo cautelar, sobre todo dirigido a ponderar el rol y la relevancia del mencionado debate de cara a la creciente atención que el ED está hoy despertando.

1. Sistemas dinámicos y explicación científico-cognitiva

El ED comprende fundamentalmente aquellos abordajes que se asientan de modo central en el uso de la teoría matemática de sistemas dinámicos aplicada al estudio de los procesos cognitivos. Un número de filósofos y científicos articularon, de diverso modo, la idea de que esta aplicación extensiva redunda en un abordaje característico al modelado y el análisis de los datos experimentales. Puede en este sentido constatarse en el ámbito de interés un uso creciente de nociones, métodos analíticos y herramientas gráficas provistas por dicha teoría que permite conceptualizar y modelar sistemas complejos en su evolución temporal. En términos generales, aquella es usada a los fines de guiar de diferentes maneras la investigación en áreas dispares de las ciencias cognitivas.

La aplicación de los conceptos y métodos de la teoría de sistemas dinámicos permite abordar los fenómenos cognitivos en términos de trayectorias en un espacio de estados posibles de un sistema previamente definido – generalmente a través de sistemas de ecuaciones diferenciales –, y estudiar el curso de su desarrollo temporal. Los modelos obtenidos tienden a caracterizar el comportamiento del sistema de interés en términos de variables de orden superior que describen estados globales del mismo, lo cual les confiere un rico poder descriptivo. En particular, mientras por una parte la noción de representación interna o mental constituye un constructo teórico crucial para cierto tipo estándar de abordaje en las ciencias cognitivas, por otra parte el foco explicativo dinamicista está puesto en la estructura del espacio de posibles trayectorias del sistema estudiado y las fuerzas internas y externas que las moldean. Esto vale para tres tipos de modelos matemáticos comunes en la investigación reciente: modelos cualitativos, modelos cuantitativos (distinguibles centralmente en términos de la precisión de las predicciones que permiten) y el caso particular de las redes neuronales.

Cabe destacar la amplitud potencial de la aplicación de este estilo de modelado. El lenguaje dinamicista habilita la construcción de modelos que integran aspectos dispares del particular fenómeno bajo estudio, con un énfasis puesto en el despliegue temporal de los mismos, sobre diferentes escalas temporales. De modo crucial, esto posibilita hacer operativo, en el contexto de estudio de los más diversos fenómenos y bajo cualesquier ulterior restricción metodológica (relativa a áreas puntuales dentro de las ciencias cognitivas), la consideración inclusiva de los aspectos pertenecientes al sistema cognitivo (definible en términos neuronales, cognitivos o conductuales)

así como de los aspectos propios del cuerpo del agente y del medio ambiente donde éste se desempeña[1].

Ahora, a pesar de su desarrollo continuado, la recepción generalizada frente al ED ha sido en buena parte negativa, tanto en sus años inaugurales hacia mediados de la década del '90 como, aunque en menor medida, durante los últimos años. Más recientemente, el de las credenciales explicativas del ED ha sido en particular un frente de ataque común por parte de filósofos y comentadores. En la literatura, pueden distinguirse dos estrategias críticas, diferentes pero complementarias, que se han utilizado en esta dirección.

Uno de los ataques más contundentes por parte de estos críticos (Rosenbaum 1998, Dietrich & Markman 2001, Bechtel 2001, van Leeuwen 2005) ha sido el de negar de raíz, debido a motivos inherentes a la misma utilización de la teoría de sistemas dinámicos, el carácter explicativo de las investigaciones dinamicistas. De acuerdo con esto, el lenguaje matemático de la teoría de sistemas dinámicos, esto es, el corazón operativo del ED, no llega a ser más que una buena herramienta descriptiva, no explicativa, para el comportamiento de sistemas complejos.

Rosenbaum (1998), por ejemplo, pone esta crítica en términos de las ecuaciones no lineales usadas para la construcción de los modelos. La idea es que, si los términos de las ecuaciones son referencialmente opacos – hasta tanto las variables utilizadas describan estados globales del sistema modelado –, todo lo que las propuestas dinamicistas pueden ofrecer son sofisticadas redescripciones, pero nunca explicaciones, de los fenómenos tratados. En esta misma dirección, van Leeuwen escribe: "es aún poco claro cómo este particular uso, corporizado y no representacional, de la teoría de sistemas dinámicos *por sí solo* […] podría generar explicaciones psicológicas y no meras descripciones del comportamiento" (van Leeuwen 2005, 286; mi traducción y subrayado) y continúa más adelante: "esto es, el modelo no ofrece elucidación del mecanismo subyacente al cambio de estado" (van Leeuwen 2005, 288; mi traducción). La noción de mecanismo es la protagonista de la segunda oleada de embestidas contra el potencial explicativo del ED.

Recientemente, ha crecido un debate, que ha tomado peso propio en el seno de la comunidad filosófica, en torno de la noción de mecanismo como basilar para la explicación (por ejemplo, Glennan 2002, Craver 2006, Bechtel 2008) y, en particular, de su relación con el ED (Chemero & Silberstein 2008, Kaplan & Bechtel 2011, Zednik 2011, Kaplan & Craver 2011). En este caso, el problema reside en el alcance de los modelos mecanicistas en las ciencias cognitivas y la

1 Es interesante ver este rasgo del ED, esto es su enorme aplicabilidad, como aquel que de algún modo conecta las líneas de trabajo caracterizadas por la aplicación de la teoría de sistemas dinámicos con la muy ventilada recuperación reciente de la contribución del cuerpo y del medio ambiente en los procesos cognitivos (ver por ejemplo Clark 1999).

compatibilidad entre aquellos y los modelos dinamicistas. Ambos ataques se complementan en la medida en que, por un lado, se decreta la insuficiencia del enfoque entendido como abordaje científico-cognitivo autónomo y, por otro lado, se lo hace subsidiario de estrategias orientadas hacia la búsqueda de mecanismos cognitivos.

En la medida en que el ED constituye un estilo de abordaje característico que excede la investigación estrictamente neurocientífica, donde ha sido primeramente explorado – haciéndose fuerte también en áreas como la teoría de la percepción, la robótica cognitiva o la psicología del desarrollo –, es menester hacer una aclaración en torno de la independencia relativa entre la cuestión del formato explicativo, por un lado, y las problemáticas en torno del reduccionismo, por otro.

Puede decirse que, en mayor o menor medida, ambas críticas esgrimidas tienen como trasfondo una determinada tradición filosófica en torno de la explicación científico-cognitiva, desde la que luego se desprenden posiciones específicas a las que les atañe la cuestión del reduccionismo. En términos generales, de acuerdo con esta tradición, una explicación en ciencias cognitivas daría cuenta del comportamiento de un sistema en términos de la interacción causal de sus partes componentes y de su contribución relativa al procesamiento de la información necesario para la tarea bajo estudio. Un desarrollo representativo en esta dirección ha sido el de enmarcar este tipo de explicación dentro de la heurística doble de la descomposición y la localización, especialmente a partir del tratamiento ofrecido en el ya clásico Bechtel y Richardson (1993). Lo relevante de esto para ponderar en particular el impacto del ED en el campo[2] es que brinda un espacio común para cobijar posiciones contrapuestas en lo que respecta a la autonomía de niveles psicológicos de teorización y el problema relacionado del reduccionismo para las ciencias cognitivas.

El punto central es que el problema de explicar fenómenos como los cognitivos comprenderá dos actividades, por lo menos analíticamente distinguibles: la descomposición del proceso cognitivo en términos de las estructuras, subprocesos y operaciones necesarias que trabajan de manera coordinada para desempeñar alguna actividad (lo que Bechtel y Richardson denominan descomposición estructural y funcional) y la subsiguiente localización en el cerebro de las estructuras específicas responsables de esas actividades (que los mismos autores entienden en términos de un mapeo entre actividades y operaciones, por un lado, y los componentes estructurales asociados, por otro).

2 Como lo argumentan Chemero y Silberstein (2008), el problema para el ED es el de si sus explicaciones deben entenderse como mecanicistas – en esta versión aplicable al grueso de las ciencias cognitivas –, mientras recién una respuesta positiva a esta cuestión nos llevaría al problema del reduccionismo. Lo que aquel problema supone en este sentido es meramente que puede darse cuenta del comportamiento de un sistema en términos de las funciones desempeñadas por sus partes y la organización de su interacción.

Lo interesante para el caso a mano es que este esquema es tan general que puede englobar una enorme diversidad (1) en los niveles en que se atacan los problemas – por ejemplo, en términos de abordajes tanto *top-down* (esto es, que toman como punto de partida las capacidades del sistema) como *bottom-up* (que toman como punto de partida los componentes estructurales del mismo) –, (2) en las herramientas experimentales, de observación y de modelado privilegiadas – desde diversas técnicas de neuroimagen hasta métodos restringidos al análisis de las tareas como los llamados análisis funcionales (Cummins 1975) o bien métodos inferenciales indirectos (Glennan 2005) como el uso de tiempos de reacción – así como (3) en la clase de dispositivos explicativos elaborados (modelos de diverso tipo y función, elaboraciones teóricas, aparatos matemáticos y simulaciones computacionales). La descomposición del sistema, en términos de subprocesos y operaciones que trabajan de manera coordinada, asociada con su localización en el cerebro es en este sentido un pilar fuerte en la interpretación de la investigación científico-cognitiva reciente.

2. Algunas consideraciones cautelares

Ha sido defendido un número de posiciones encontradas frente a este debate reciente con atención a la irrupción del ED en el panorama contemporáneo de investigación – posiciones que defienden (1) cierta novedad inherente al enfoque en términos de su potencial descriptivo y predictivo (por ejemplo, Giunti 1997, van Gelder 1998, Stepp, Chemero & Turvey 2011), (2) un tipo de explicación en el sentido hempeliano de cobertura legal (Walmsley 2008) y (3) los modelos dinamicistas como herramientas descriptivas para representar cómo ciertos mecanismos complejos trabajan y, de este modo, como parte del proyecto de proveer explicaciones mecanicistas (Kaplan & Craver 2011, Zednik 2011, Kaplan & Bechtel 2011). Aunque no puedo detenerme aquí en aspectos específicos de estas posiciones (ver Venturelli 2012), quiero a continuación adelantar algunas consideraciones cautelares generales sobre el desarrollo del debate con alguna sugerencia para su redireccionamiento.

En la medida en que la problemática que se ha instalado en torno del ED constituye una aplicación de debates pertenecientes a la filosofía general de la ciencia al aún joven campo de las ciencias cognitivas, hereda algunos sesgos problemáticos que merecen ser destacados. Koertge (1992), en una incisiva revisión, propone preguntarse por qué tipo de problemas debiera de hecho resolver una teoría filosófica de la explicación. Con atención a esto, estimo que mirar al problema desde un contexto de indagación lo redefiniría provechosamente, especialmente en la dirección de desestimar preguntas del tipo: ¿este modelo particular explica o no?

En particular, tratándose de un desarrollo muy reciente en el estilo de trabajo en las ciencias cognitivas, parece adecuado intentar evitar la tendencia a encasillar resultados específicos en tipos determinados de explicación – poniendo de este modo entre paréntesis la cuestión preliminar de la distintividad del abordaje dinamicista al modelado, en términos de lo que ofrece a nivel explicativo. Puede resultar según esto prematuro y quizás contraproducente asociar la investigación asentada sobre herramientas de sistemas dinámicos con una concepción filosófica particular de la explicación como puede ser el modelo hempeliano (Walmsley 2008) o el mecanicista (Zednik 2011).

Por una parte, impacta como un requisito muy estricto el de apelar a leyes en algún punto del esfuerzo explicativo: este requisito no es connatural a la construcción de modelos dinámicos vía sistemas de ecuaciones ni a su análisis mediante la teoría de sistemas dinámicos (excepciones prominentes son las de Thelen et al. 2001 y Beer 2003). Más allá de los múltiples problemas aducidos en contra del modelo hempeliano (ver a este respecto Kaplan & Craver 2011), entiendo que su aplicación al ED constituye un salto arriesgado y, en especial, innecesario para circunscribir el estilo de trabajo dinamicista en términos de su valor explicativo.

Por otra parte, el requisito crecientemente invocado (Bechtel 2001, Kaplan & Craver 2011) de un acercamiento a los abordajes mecanicistas (por ejemplo, en la medida en que pueden contribuir a la heurística descomposición / localización) prejuzga y restringe innecesariamente futuras dinámicas de cambio teórico en la dirección de la búsqueda de mecanismos. En la medida en que este tipo de restricciones estén montadas sobre requerimientos metodológicos reduccionistas, fracasan en dar cuenta del rol de los modelos dinamicistas en diversas áreas en las que han sido exitosamente aplicados (como, nuevamente, la psicología del desarrollo o la robótica cognitiva). Bajo esta luz, se avizora una perspectiva restrictiva de la investigación dinamicista, entendida como subsidiaria de estrategias dirigidas a identificar y caracterizar mecanismos cognitivos.

Cabría, por otra parte, estimular cierta cautela respecto de la idea de que existe un tipo de propiedad o relación (relevancia causal, unificación, etcétera) que captura el concepto de explicación a lo largo de diversas aplicaciones locales. Koertge afirma:

> "Podría ser útil clasificar las exploraciones científicas de acuerdo con su estructura lógica y epistemológica. Resta por verse, sin embargo, si las preguntas que normalmente consideraríamos como pedidos de explicación caen dentro de una única celda de la tipología del problema." (Koertge 1992, 97; mi traducción)

En términos más generales, esta suerte de consideraciones se enmarca en el seno de una denuncia respecto del papel preponderante que suele asumir la explicación en desmedro de otros

objetivos cognitivos. Para el caso del ED, puede destacarse su habilidad para formular nuevas preguntas, divisar modos para responderlas y refinar viejas respuestas, incluso en la dirección de abordar nuevos estándares de precisión. Una habilidad de este tipo podría ponderarse, por ejemplo, en términos de la medida en que el ED ofrece nuevas perspectivas sobre el rol de la estructura temporal en el comportamiento y la cognición.

Cabe finalmente observar que es la tendencia a reducir el ED a (o sesgarlo hacia) la obtención de explicaciones que habilita el sentido descalificador de ciertas posiciones en el debate: se presupone en este sentido el punto general de que la evaluación de modelos se da en términos de sus credenciales explicativas. Esto muestra además cómo el rol de buena parte del debate en relación con las disciplinas asociadas se encuentra netamente inserto en un contexto de justificación. Una mirada más procesual podría, en cambio, exhibir múltiples rasgos de los modelos, generalmente desestimados por parte de la comunidad filosófica, rasgos como su poder predictivo, su papel como generadores de hipótesis, el control sobre si y cómo se produce determinado fenómeno, la descripción de patrones complejos de comportamiento, la identificación de principios formativos de estos patrones, etcétera.

Conclusión

He intentado sugerir que el debate revisado en torno del estatus explicativo del ED es parasitario de una visión hermética de los emprendimientos científico-cognitivos así como de algún supuesto unitarista respecto de un modelo privilegiado de explicación aplicables a estilos diferentes de búsqueda. Favorecí por el contrario una mayor ponderación de la relación entre explicación y otros objetivos cognitivos, y una mayor sensibilidad hacia los factores contextuales asociados: una tendencia de este tipo cobra especial interés considerando el extraordinario número de asunciones de carácter teórico y metodológico subyacentes a cada línea individual de trabajo en las ciencias cognitivas – asunciones que repercuten en el recorte de los problemas y los niveles en que se atacan (molecular, celular, de sistemas neuronales, conductual, interpersonal, social, evolutivo, etcétera), las herramientas experimentales y de observación adoptadas, diversos tipos de modelos, lenguajes matemáticos, simulaciones computacionales y elaboraciones teóricas así como en las complejas relaciones entre estos.

REFERENCIAS

Bechtel, W. 2001, ‘The compatibility of complex systems and reduction’, *Minds and Machines* 11, 483-502.

Bechtel, W. 2008, *Mental Mechanisms*, London, Routledge.

Bechtel, W. & Richardson, R.C. 1993, Discovering Complexity, Princeton, Princeton University Press.
Beer, R. 2000, 'Dynamical approaches in cognitive science', *Trends in Cognitive Sciences* 4, 91-9.
Beer, R. 2003, 'The dynamics of active categorical perception in an evolved model agent', *Adaptive Behavior* 11, 209-43.
Chemero, A. & Silberstein, M. 2008, 'After the philosophy of mind', *Philosophy of Science* 75, 1-27.
Clark, A. 1999, 'An embodied cognitive science?', *Trends in Cognitive Sciences* 3, 345-51.
Craver, C. 2006, 'What mechanistic models explain', *Synthese* 153, 355-76.
Cummins, R. 1975, 'Functional analysis', *Journal of Philosophy* 72, 741-64.
Dietrich, E. & Markman, A. 2001, 'Dynamical description versus dynamical modeling', *Trends in Cognitive Sciences* 5, 332.
Giunti, M. 1997, *Computation, Dynamics, and Cognition*, New York, Oxford University Press.
Glennan, S. 2002, 'Rethinking mechanistic explanation', *Philosophy of Science* 69, 342-53.
Glennan, S. 2005, 'Modeling mechanisms', *Studies in History and Philosophy of Biological and Biomedical Sciences* 36: 443-64.
Kaplan, D. & Bechtel, W. 2011, 'Dynamical models', *Topics in Cognitive Science* 3, 438-44.
Kaplan, M. & Craver, C. 2011, 'The explanatory force of dynamical and mathematical models in neuroscience', *Philosophy of Science* 78, 601-27.
Koertge, N. 1992, 'Explanation and its problems', *British Journal for the Philosophy of Science* 43, 85-98.
Rosenbaum, D. 1998, 'Is dynamical systems modeling just curve fitting?', *Motor Control* 2, 101-04.
Stepp, N., Chemero, A. & Turvey, M. 2011, 'Philosophy for the rest of cognitive science', *Topics in Cognitive Science* 3, 425-37.
Strogatz, S. 1994, *Nonlinear Dynamics and Chaos*, Reading, Addison-Wesley.
Thelen, E., Schöner, G., Scheier, C. & Smith, L. 2001, 'The dynamics of embodiment', *Behavioral and Brain Sciences* 24, 1-86.
van Gelder, T. 1998, 'The dynamical hypothesis in cognitive science', *Behavioral and Brain Sciences* 21, 615-65.
van Gelder, T. & Port, R. 1995, 'It's about time', in Port, R. & van Gelder, T. (eds.), *Mind as Motion*, Cambridge, MIT Press, 1-43.
van Leeuwen, M. 2005, 'Questions for the dynamicist', *Minds and Machines* 15, 271-333.
Venturelli, N. 2012, '¿Puede hablarse de una explicación dinamicista en las ciencias cognitivas?', *Ludus Vitalis* XX, in press.
Walmsley, J. 2008, 'Explanation in dynamical cognitive science', *Minds and Machines* 18, 331-48.
Zednik, C. 2011, 'The nature of dynamical explanation', *Philosophy of Science* 78, 236-63.

INDISPENSABILITY AND MATHEMATICAL EXPLANATION

SUSAN VINEBERG
Wayne State University
susan.vineberg@wayne.edu

ABSTRACT: Recently, an 'enhanced' version of the Indispensability Argument has been advanced which claims that we ought to accept the existence of mathematical objects because of the indispensable explanatory role that they play in science (Baker 2009). In particular, Baker argues that there are mathematical explanations of physical facts that are accepted in science, which requires accepting the mathematical objects that figure in these explanations. Baker identifies the challenge for the nominalist as that of specifying "the distinctive way in which ordinary material bodies can play an explanatory role…so that it can be said that abstracta which are mentioned in the context of scientific explanations are not explanatory in that way". This paper addresses this challenge.
Keywords: Explanation, Indispensability, Mathematics, Realism

1. The Enhanced Indispensability Argument

Many versions of the Indispensability Argument have been given for the claim that the use of mathematics in science commits us to mathematical objects, involving a variety of assumptions, but here I focus on the version Baker gives as follows:

(1) We ought rationally to believe in the existence of any entity that plays an indispensable explanatory role in our best scientific theories.

(2) Mathematical objects play an indispensable explanatory role in science.

(3) Hence, we ought rationally to believe in the existence of mathematical objects.

Baker takes (1) to be a basic tenet of scientific realism. He defends (2) with the following purported mathematical explanation of why certain species have prime life-cycles:

E1. Having a life-cycle period that minimizes intersection with other (nearby/lower) periods is evolutionarily advantageous.

E2. Prime periods minimize intersection.

E3. Hence, organisms with periodic life-cycles are likely to evolve periods that are prime.

E4. Cicadas in ecosystem-type E are limited by biological constraints to periods from 14 to 18 years.

E5. Hence, cicadas in ecosystem-type E are likely to evolve 17-year periods.

Thus, explaining that

E*. Cicada in ecosystem-type E have a life-cycle of 17 years.

Baker's indispensability argument is valid, and so its legitimacy rests on whether the explanation above, or something similar, is a genuine mathematical explanation of a physical fact, and on whether the first premise is true.

2. Some Responses

One might say that Baker's example does not involve an indispensable use of mathematics in explaining a physical fact. Perhaps mathematics isn't properly part of the explanation here, or essentially the same explanation could be written so as to avoid reference to mathematical objects. The latter response has been much discussed,[1] and as Baker rejects both, I won't consider them here. Instead of showing that Baker's explanation is not committed to mathematical objects, Saatsi (Saatsi 2011) argues that mathematical claims about primes are inessential to explaining E, by giving an alternative explanation that replaces Baker's E2-E3 with

ES. For periods in the range 14-18 years the intersection minimizing period is 17.

This yields an explanation that he claims is nominalistically acceptable.

1 See, (Burgess and Rosen 1997).

Baker responds that his explanation is better for being more general; moreover, it accounts for the connection between prime numbers and intersection minimizing periods, which generalizations of Saatsi's argument cannot. Thus, whether or not Saatsi has shown that mathematics is dispensable in giving an explanation of E*, Baker holds that it is nonetheless indispensable to the best explanation, and since scientific realism holds that we are to accept inference to the best explanation (IBE), we are to accept his explanation of E*, and the accompanying mathematical claims, which carry with them a commitment to abstract mathematical objects. Saatsi disputes the assertion that mathematical facts about primes contribute to the explanation of E*, and regards the general the role of mathematics in science as representational and computational, rather than explanatory. Nevertheless, Baker's claim that his is the better explanation has some appeal, and this requires some response.

3. Inferences from Explanations

Scientific realism typically assumes that some form of IBE is legitimate. To avoid the worry that the best explanation might be the best of a very shaky set of candidates, advocates of IBE insist that it applies to explanations that meet some threshold of acceptability, but as with what counts as best, this condition is not well specified. This makes it difficult to definitively refute either the claim that Baker's explanation is the best available or that IBE is justified in this case. That said, even if IBE (or a form of IBE) is properly applied in science, there is reason to think that not all explanations, even those that meet certain criteria for being the best, warrant inference to their truth. Generality and simplicity, which are often claimed as virtues, at least in some types of explanation, need not be indicators of truth, and it is unclear that they carry the sort of ontological import in science that the advocate of the Indispensability Argument requires. For instance, Cartwright argues that IBE is not generally licensed, but maintains that scientists do infer to the best casual explanation (Cartwright 1983). Boyd also emphasizes that it is our best causal explanations of the phenomena that science is committed to accepting (Boyd 1984). If scientific realism requires only a fairly limited version of IBE, then it unclear that it can be used to argue for mathematical realism as Baker envisions. Certainly the argument will not go through with a restriction of the principle to just causal explanations, since mathematical explanations are not causal explanations. Still, perhaps a broader form of IBE can be identified, to which the putative mathematical explanations of physical facts conform, which sanctions inference to the truth of the explanation.

The Enhanced Indispensability Argument requires that there is a form of explanation, to which the putative cases of mathematical explanation conform, in which the broad grounds for

accepting explanations of this type are those that the scientific realist accepts. Of the limited accounts of mathematical explanation that have been proposed, the most promising initially for the mathematical realist's purposes seems to be Kitcher's unification account,[2] because it covers both scientific and mathematical explanation, making it potentially suitable for giving mathematical explanations of physical facts. In addition, it associates explanation with generality, which is one of the features that Baker highlights in his example. But Kitcher's account, which takes explanation to consist in deriving the phenomena from a minimal stock of argument patterns, thereby achieving a measure of unification, does not clearly support the kind of inference that the mathematical realist requires. For while there are cases, most notably Newton's Theory of Gravitation, that are regarded as receiving support on account of their unifying power, there is little consensus in support, even among scientific realists, that unification in Kitcher's sense is linked with acceptance and truth. The issue here is not so much with Kitcher's theory per se, but with the idea that simple accommodation of a theory within a theoretical framework, even when it possesses various theoretical virtues, such as generality, simplicity or unification, is a clear marker of truth.

4. Causal Explanation and an Analogue for Mathematics

Instead of proceeding from one of the few developed theories of mathematical explanation and attempting to argue that inference in accordance with mathematical explanation in that sense is warranted, an alternative approach involves extending the causal theory of explanation to cover mathematics. Causal explanation, unlike other forms based on considerations of simplicity, unification, and the like, has the feature that full acceptance of such an explanation demands accepting the causes that figure therein. There remains the question of why one should accept a given causal explanation. This is a difficult matter, but given the starting point of scientific realism I put it aside. The motivating thought here is that a proper analogue of causal explanation for mathematical explanation offers the best hope for grounding a commitment to the mathematics that figures in an explanation. Of course whether Baker's example would conform to such a model, and whether, in the end, this would provide a reason for accepting mathematical objects, remains to be seen. But the approach, which seeks to associate explanation with locating conditions of dependence,[3] holds some promise here, particularly since there is a strain of mathematical research that appears to aim at just this.

2 See, (Kitcher 1989)
3 This differs from Kitcher's view on which dependence is to be cast in terms of the argument patterns that constitute explanations.

In trying to expand the causal model of explanation to cover mathematics, the central idea of Woodward's theory of causal explanation is useful, namely that such explanations involve citing those factors that make a difference to whether a given phenomena occur. [4] In causal explanations proper, such factors are those that can potentially be intervened upon to change whether an event happens. This condition is inapplicable to mathematics, and thus mathematical explanation cannot be causal explanation; nevertheless, sense can be made of a mathematical claim depending upon another.

A causal explanation of some phenomena is one that details, relative to a background set of assumptions, the conditions under which it occurs, and without which it does not. An analogous mathematical explanation of a fact or theorem would spell out, in terms of some background mathematical structure, a set of mathematical conditions under which the fact or theorem holds, and without which it does not. Such conditions then give rise to a notion of dependence cast in terms of a given background structure. Establishing such dependence involves determining what follows from a given set of axioms, but insofar as the axioms themselves are not purely logical, the dependence established will be of a more general structural sort. The concept of mathematical explanation here is in sympathy with the theory of explanatory proof offered by Steiner.[5] On his account an explanatory proof of a theorem is one that

> makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends upon the property. It must be evident, that is, that if we substitute in the proof a different object of the same domain the proof collapses; more, we should be able to see as we vary the object how the theorem changes in response (Steiner 1978, 143).

Although Steiner's theory addresses explanatory proofs, the idea that such proofs involve citing properties on which the result holds is central to the analogue of causal explanation suggested here. To explain, as with causal explanation proper, involves citing those factors that make a difference, rather than simply deriving the phenomena from a set of principles or assimilating it to some argument form.

Although there is much to be detailed about this form of mathematical explanation, here I'll note just a simple case to illustrate the view and its contrast with the unification view. Given some theorem of analysis we might explain it on the unification theory in set theory by giving a set theoretic definition (representation) of the concepts involved therein, and then deduce the theorem

4 (Woodward 2003)
5 (Steiner 1978).

from the axioms of ZF, or perhaps some extension thereof, although one might well question whether doing so generally offers much understanding. Unlike the conception of explanation that undergirds the unification view, the analogue of causal explanation for mathematics does not take explanation to consist in its assimilation to a tight collection of principles or axioms. Rather explanation in set theory on this model would consist in finding the minimal set theoretic assumptions required for the theorem to hold. Locating such set theoretic conditions for ordinary (non- set theoretical) theorems is the aim of Friedman's program of Reverse Mathematics.[6] The mathematical analogue to causal explanation thus characterizes this as an explanatory project of a very different sort than that embraced by the unification theory. It also goes a long way towards undergirding a naturalistic motivation for a large current of other foundational research, which aims at locating minimal conditions for theorems to hold, and accordingly the establishment of strong results, rather than identifying strong hypotheses from which much can be proven.

With this sketch of a different sort of explanantory model for mathematics, the question arises about the ontological implications of determining some relation of mathematical dependence. Consider for instance the claim that a particular theorem of ordinary mathematics depends upon some (controversial) set theoretic axiom. I take it that this provides a powerful reason for accepting the axiom, indeed one that is generally more convincing than establishing that it enables the derivation of a variety of attractive theorems. Of course, this will only justify taking the objects mentioned in the axiom as real if the original theorem is similarly so read, so it is unclear that mathematical realism must go along with scientific realism. In any event, the upshot of the next section is that Baker's explanation has not been shown to depend upon the existence of mathematical objects.

5. Baker's Example and Mathematical Dependence

The question is whether Baker's explanation warrants acceptance of a mathematical theory. If we cite the fact that 17 is prime in answer to why the cicadas in a particular region have 17 year life-cycles, then the explanation isn't entirely satisfactory, since it is not being prime per se that accounts for 17 as the optimal predator-avoiding life-span, but that in the range of 14-18 years, this is the life-span that minimizes intersection.[7] This point reinforces the idea that the example proceeds by assimilating the fact about the life-cycle length to a theory that incorporates

6 See (Harrington et al. 1985).
7 For a given possible time period in years, the intersection minimizing life-span for species in that interval is obtained by eliminating those that are not relatively prime, and then picking the largest number that remains.

mathematical facts about primes, rather than establishing a dependence on those facts. This follows along with Saatsi's alternative explanation, which highlights the idea that it is the minimization of intersection with other species within the given constraint on the life-cycles that does the crucial explanatory work.

Baker is committed to the claim that the theory of prime factorization provides more than simply an elegant means of characterizing those life-cycles that minimize intersection; it also serves an additional explanatory function. This seems unclear, but insofar as it is explanatory, the question is whether this supports the inference to the mathematical objects figuring in the theory. In order to support the inference along the lines that IBE is clearly employed by scientific realists, the right sort of dependence must be established between the objects of this further explanatory mathematical theory and the minimization of intersection between species. One could take the dependence to reside in the systematization of the theory, but this looks insufficient to require accepting the existence of the mathematical objects of the theory. I've suggested instead that the best case for grounding acceptance of theory in explanation is when the explanation cites a set of difference makers for the explanandum.

In the cicada example, mathematical objects are used to describe certain structural features of the life-cycles, with the key feature that is evolutionarily advantageous characterized in terms of the divisors of the period of the life-cycle. But granting that those periods that minimize intersection are characterized by certain number theoretic facts, it has not been shown that the explanadum here depends upon the existence of numbers. It is the particular structural features that play the difference making role, and not the mathematical objects (numbers) that are referenced in exhibiting the structure of those life-cycles. One could say that mathematical facts about primes are part of the explanation of why prime life-cycles are evolutionarily advantageous, because those facts characterize the relevant difference makers in the structural features of various possible life-cycles. But it is not the mathematical objects themselves that are the difference makers; they are simply used in characterizing those (possible) structures whose relational features are the difference makers. In this respect, the case is strikingly different from the cases of causal explanation in which it is shown how the existence of certain physical objects or presence of physical properties make a difference as to whether a given phenomenon occurs.

Mathematical realists will surely insist that any talk of mathematical structures here commits us to abstract objects. But that requires a different argument, for it is hard to see that there is anything special about the scientific example that prohibits taking the traditional nominalist stance that structures do not exist independently of the objects in which they occur. On this nominalist

picture, mathematics shapes the explanation, but its fundamental role is representational, in accordance with Saatsi's view. Mathematics characterizes the structures of the life-cycles, with the mathematical difference makers in turn distinguishing those features that are the physical difference makers. This reveals the answer to the question with which we began, namely that abstracta do not play the sort of difference making role as physical objects, and thus do not have the same explanatory role. To the extent that we say here that the mathematics is nonetheless to be counted as having an indispensable explanatory role, the Enhanced Indispensability Argument fails because its initial premise is not compelling.

REFERENCES

Baker, Alan (2009), "Mathematical Explanation in Science", *British Journal for the Philosophy of Science* 60, 611-633.

Boyd, Richard (1984), "The Current Status of Scientific Realism", in Jarrett Leplin (ed.), *Scientific Realism*, Berkeley: University of California Press, 41-82.

Burgess, John P., and Gideon Rosen (1997), *A Subject With No Object*. Oxford: Oxford University Press.

Cartwright, Nancy (1983), *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Harrington, L.A., M.D. Morley, A. Scedrov, and S.G. Simpson, eds. (1985), *Harvey Friedman's Research on the Foundtions of Mathematics*. North Holland.

Kitcher, Philip (1989), "Explanatory Structure and Causal Unification", in Philip Kitcher and Wesley C. Salmon (eds.), *Scientific Explanation*, Minneapolis: University of Minnesota Press, 410-505.

Saatsi, Juha (2011), "Discussion: The Enhanced Indispensability Argument: Representational vs. Explanatory Role of Mathematics in Science", *British Journal for the Philosophy of Science* 62, 143-154.

Steiner, Mark (1978), "Mathematical Explanation", *Philosophical Studies* 34, 135-151.

Woodward, James (2003), *Making Things Happen*: Oxford.

VERDADES SIGNIFICANTES*: O COMO INCLUIR LO POLÍTICO EN LA VERDAD

ZENAIDA YANES ABREU
Universidad de La Laguna

ABSTRACT: To accept the social character of scientific knowledge is a challenge for the Contemporary Philosophy of Science. The object of reflection changes and it is imperative that our discipline studies the goals of science and decision making process, rather than analyzes the structure of theories. The model of Well Ordered Science proposed by Philip Kitcher is an example of this new direction. The purpose of this paper is to show how the concept of significant truth can fit social and political elements with rationality.
KEY WORDS: Science, policy, truth, objectivity, significance and racionality.

1. La Ciencia Bien Ordenada. Un marco donde conjugar ciencia y política

Philip Kitcher es uno de los filósofos que ha asumido el reto de alcanzar una concepción naturalista del conocimiento, donde se conjugan los elementos racionales y epistémicos con asunciones valorativas propias de la ciencia, sin renunciar al carácter normativo propio de cualquier propuesta filosófica. Con el modelo de Ciencia Bien Ordenada, el autor pretende recuperar el ideal baconiano de búsqueda del bien común, de progreso y mejora global, como principal meta de la actividad científica; pero esto no es sólo el rescate de un ideal que guíe las prácticas de conocimiento, es la ruptura con la imagen tradicional de la ciencia donde la verdad y la neutralidad es el único objetivo; y es también, la conciencia plena de la relación entre ciencia y bien público, es decir, entre ciencia y política.

Ya hacen hincapié en esta relación entre epistemología y política otros autores, especialmente de corte sociológico, o las epistemólogas feministas, que denuncian con rotundidad que la ciencia es un instrumento de dominación que perpetúa el status quo, afirmando que cualquier transformación política debe ir acompañada de una transformación del modelo de conocimiento. Sin caer en la imagen de ciencia como producto ideológico, Helen Longino señala que *ningún proyecto de transformación política puede hacerse sin la ciencia* (Longino 1990, x) y tampoco

* El término utilizado por Philip Kitcher es *significant truth*, que podría ser traducido como verdades significativas, no obstante se ha optado por traducir como verdades significantes para remarcar la dimensión del significado.

puede hacerse sin una mejor comprensión filosófica de la investigación científica que la actualmente disponible. De acuerdo con esta idea, Kitcher reconoce la reciprocidad entre ciencia y política, y señala que:

> "(…) es importante hacer todo lo que se pueda hacer para asegurar que los descubrimientos sobre la naturaleza humana se traduzcan en una política social sensible a las necesidades de los desfavorecidos" (Kitcher 2001, 8-9).

La pregunta del para qué la ciencia abre la puerta a los valores sociales e implica aceptar que la ciencia es una práctica contextualmente interdependiente. Del mismo modo que el contexto determina el desarrollo de las prácticas científicas, éstas pueden transformar dicho contexto. Mantener esta imagen de la práctica científica al tiempo que se defienden las tesis realistas es una ardua tarea que implica inevitablemente la redefinición de conceptos como verdad, objetividad, significación y éxito.

2. Veritismo. El comienzo de la resignificación del concepto

Una aceptación dogmática de cualquier creencia resulta cuando menos pretencioso teniendo en cuenta que la historia de la ciencia nos ha dado numerosos ejemplos de la falibilidad de nuestro conocimiento. Así, ante el abuso de la noción de verdad por parte de las posturas realistas tradicionales, los nuevos realismos, pretenden prescindir de la connotación metafísica del término y apuestan por rehabilitar este concepto estableciendo una noción mínima de verdad y haciendo hincapié en el uso cotidiano del mismo.

Con el fin de liberar a la noción de verdad de la rigidez del realismo clásico e incorporar el carácter social del conocimiento, Kitcher utiliza el concepto de veritismo. Esta noción de verdad propuesta por Goldman está basada en la verdad por correspondencia en sentido aristotélico, donde la relación con la realidad es de descripción, y el elemento clave es la inferencia del éxito a la verdad.

El éxito es uno de los criterios de los que tradicionalmente se ha servido la ciencia para afirmar la verdad de sus teorías. Goldman entiende que el éxito descriptivo es el barómetro de la verdad de las teorías, mientras que Kitcher pone el énfasis en el éxito predictivo. La predicción se puede presentar también como un ejemplo de que la correlación entre éxito y verdad puede ser inválida, ya que una teoría puede ser exitosa en sus predicciones y no por ello ajustarse de forma precisa a la realidad. No obstante, aunque inferir la verdad de una práctica exitosa no es un criterio de máxima fiabilidad, sí es fiable en cierto grado. Y es en este aspecto en el que coinciden ambos autores, la nueva noción de verdad debe ser entendida como una cuestión de grado de creencia. La

rehabilitación de la noción de verdad supone abandonar nuestra categorización de la misma en función del la aceptación, suspensión o negación del juicio, y entenderla como una cuestión de grado. Cuanto más preciso, sistemático y refinado sea el éxito predictivo o descriptivo, más probabilidades hay de que la inferencia del éxito a la verdad sea correcta, y más fiable es nuestro conocimiento.

Por otro lado, la inferencia del éxito a la verdad debe ser entendida como una práctica que nos permite apoyar nuestras creencias sobre postulados de teorías que funcionan, así, estamos justificados a creer en la verdad de estas teorías, o al menos en determinadas partes. Es decir, en la medida en que la verdad se aplica a creencias y a estamentos concretos, es posible hablar de verdades parciales. Hay elementos concretos de las teorías de los que se puede afirmar su carácter de verdad sin que esto implique necesariamente afirmar la verdad de una hipótesis o teoría en su conjunto.

Otro de los aspectos más relevantes para la rehabilitación de la noción de verdad es la introducción de elementos pragmáticos y sociales en la noción de éxito. Para obtener resultados exitosos es clave realizar una predicción sistemática y refinada, pero además:

> “ (...) una ciencia es exitosa cuando proporciona significados a fines hipotéticos, fines que son numerosos, diversos y difíciles de conseguir de forma individual (Kitcher 2001, 17)”.

El carácter colectivo de la ciencia se deriva de multiplicidad y dificultad de los fines de la empresa científica, fines que no sólo son prácticos sino también epistémicos, lo que evita una interpretación meramente pragmática o instrumentalista del éxito científico.

De este modo, el veritismo flexibiliza la noción de verdad al aceptar la gradación y la parcialidad como elementos propios de la misma, y reconocer el carácter social y colectivo del éxito científico.

3. Rompiendo la rigidez de la objetividad

La objetividad es la que da fundamento a la justificación de las teorías, y ésta a su vez se basa en la solidez de las evidencias. Mantener el concepto de verdad en un realismo mínimo pasa por defender un concepto amplio de realidad y aceptar que hay múltiples elementos que sirven de evidencias para las teorías científicas, es decir, que no es posible ofrecer una descripción precisa de los elementos que garantizarían la objetividad científica. Lo cual no implica necesariamente la negación de la misma, sino una revisión de los elementos que tradicionalmente la han definido, romper con la rigidez de este concepto derivado principalmente por una visión mitificada de la

ciencia, y establecer una serie de criterios propios que nos permitan reestablecer la objetividad científica en el marco de una imagen social de la ciencia.

El éxito científico, como ya vimos, es un criterio válido para establecer la verdad de las teorías y por tanto su objetividad. Sin embargo, la tesis de la infradeterminación de las teorías por los datos nos muestra que teorías rivales puedan tener el mismo grado de éxitos ya sean superpuestos o inconexos. Kitcher no niega esta posibilidad, pero lo asume como un momento transitorio de la investigación, en ningún caso acepta que esto pueda darse de forma permanente. Una infradeterminación permanente implicaría que la primacía de una teoría u otra depende principalmente de factores no epistémicos, y aunque no se niega el rol que juegan estos en la toma de decisiones no se acepta la primacía de estos. La tesis de la infradeterminación pone sobre la mesa la importancia de los valores o factores no epistémicos en la elección de las teorías, pero no se puede asumir como una tesis definitiva y universal, ya que esta visión de la ciencia nos llevaría al constructivismo y relativismo, negando así cualquier estándar de racionalidad.

La naturaleza es independiente de nosotros, pero no lo es el acto mismo de conocer. La naturaleza o la realidad es la piedra de mármol o el barro, y el científico, el artesano que moldea, da forma y marca las fronteras del material que tiene entre manos. Un mismo objeto puede ser representado de distinta forma por distintos artesanos porque cada uno tiene su propio lenguaje. Nuestras clasificaciones no construyen el mundo, pero construyen nuestra representación de la misma. No obstante, la posibilidad de alcanzar la objetividad científica se mantiene al negar la tesis de la infradeterminación de las teorías por los datos como una tesis permanente y al afirmar que hay lenguajes más apropiados que otros, y por lo tanto, representaciones más adecuadas que otras.

4. La significación. Un matiz que resignifica la verdad y la objetividad

El objetivo último del realismo, tal y como sabemos, es la búsqueda de la verdad, pero Kitcher añade un matiz importante, no es la búsqueda de cualquier verdad o de una verdad absoluta, es la búsqueda de verdades significantes. La concepción de la significación teórica o epistémica en Kitcher es uno de los elementos más relevantes de su propuesta, ya que aquí se ve como la significación no viene dada de forma unidireccional sino que entran en juego muchas variables.

La visión tradicional de la ciencia nos ofrece una descripción de la significación que viene determinada por la relación del contenido de las teorías con la estructura del mundo, relación en la que no intervienen elementos contextuales y cuyo principio es ofrecer entendimiento objetivo a través de las explicaciones. Desde esta óptica, las verdades epistemológicamente significativas son aquellas que se dan en repetidas ocasiones en una completa narración causal, haciendo caso omiso

de la evidencia de que en numerosas ocasiones los intereses epistemológicos e intereses prácticos están interrelacionados. Sin embargo, la aproximación de 'explicación objetiva' (la cual se identifica con verdad significante) que ofrece Kitcher es muy diferente a la visión tradicional, ésta se puede definir como: "(...) aquel conocimiento por el cual aumenta la posibilidad de que una persona alcance sus metas." (Kitcher 2001, 65), donde las metas son proyectos tanto teóricos como prácticos.

Una explicación objetiva es un conjunto de verdades que deben estar en una relación apropiada con un asunto de interés para el sujeto cognitivo, es una relación epistémicamente relevante aunque también muestran como el sujeto atiende más a unas propiedades que a otras en función de sus intereses. Hay cuestiones que nos resultan interesantes (amplían nuestro conocimiento) y otras que tienen una finalidad práctica, pero en general nuestros intereses epistémicos y prácticos están entrelazados. En este punto, Kitcher reconoce que los valores morales y sociales son intrínsecos a la práctica científica y que esto debe reflejarse en la noción de significación epistémica.

La significación epistémica contextualmente dependiente y con un papel determinante de las metas científicas, es decir de la pregunta por el para qué, rompe con la visión unificada de la ciencia e implica la aceptación del modelo de ciencia como un patchwork presentado por Nancy Cartwright. Esta visión de la práctica científica permite reflejar la multiplicidad de elementos que intervienen en la asignación de significado. La significación tiene su fuente en múltiples asuntos específicos, de modo que las conexiones que confieren significación a un asunto pueden derivarse desde distintas direcciones.

Para representar esta visión del proceso de significación tenemos los diagramas de significación. Estos gráficos nos permiten ver cuales son los distintos elementos que le dan significado a un proyecto, y como se relacionan las distintas preocupaciones y objetivos estableciendo unas u otras conexiones en función de las prioridades. En el diagrama están reflejadas numerosas cuestiones, se muestra la herencia de la significación epistémica, de dónde proviene ésta y cómo puede evolucionar en la medida en que se adquiere más información, cómo se forman nuevas conexiones, surgen nuevas dudas, etc. Los diagramas de significación rompen con la idea de que la significación epistémica sea el establecimiento o descubrimiento de una ley universal e introducen carácter pragmático a la significación. En esa red que otorga significado a una cuestión, los elementos prácticos y epistémicos están entrelazados de tal manera que un concepto teórico de la ciencia actual puede adquirir su significado en función de cuestiones que en el paradigma anterior tenían un significado eminentemente práctico. El desarrollo de la ciencia es completamente

contingente en la medida en que elecciones alternativas puede derivar la investigación en otra línea. Las investigaciones están en continua evolución, podemos variar las clasificaciones, las representaciones, etc., en la medida en que cambian nuestras necesidades y el objeto de nuestra curiosidad. Toda actividad implica transformación, y el conocer es un acto que no permanece estático, el entendimiento se va "dibujando" en base a decisiones del pasado y del presente.

Esta imagen de la significación nos lleva también a romper con la distinción entre ciencia pura y tecnología, división que se ha utilizado generalmente para articular la defensa de una ciencia neutral y objetiva. Los diagramas de significación muestran como esa distinción no es posible dado que los elementos se entremezclan inexorablemente. Evaluar una práctica científica implica evaluar todas las conexiones con proyectos pasados, presentes y futuros, y en esa red se encuentran inmersos intereses teóricos y prácticos que impiden que hablemos de ciencia pura.

Conclusión. La significación abre la puerta a los elementos sociales

El veritismo y los elementos que introduce, tales como la aceptación de la verdad como una función de grado de creencia, la parcialidad, el interés del sujeto y el carácter colectivo de la ciencia, ayudan a la descripción naturalizada del conocimiento donde se reconoce el carácter social del mismo. No obstante, la rehabilitación del concepto de verdad desde una noción mínima, no es suficiente para reinterpretar la ciencia y sus representaciones teóricas a la luz de los elementos sociales que la configuran. La noción mínima de verdad no es capaz de determinar la primacía de una verdad sobre otra desde el punto de vista epistémico, menos aún desde una perspectiva práctica. De ahí, la importancia de la significación. Las verdades significantes no atienden sólo a elementos epistémicos sino también a intereses prácticos, ambos elementos determinan la evolución de la investigación, y ambos determinan la relevancia de la misma. La significación científica propuesta por Kitcher rompe por fin con la distinción entre ciencia y tecnología, con la imagen mitificada de la ciencia pura y abstracta, y reconoce la importancia de las metas, los intereses, en definitiva los elementos sociales de la ciencia. Sin embargo, el concepto de significación descrito en *Science, Truth and Democracy* ha sido criticado ya que los valores prácticos y epistémicos se definen en función de los principios de curiosidad natural y racionalidad instrumental, introduciendo los valores sociales en un último paso del proceso de producción del conocimiento. Pero dada la relevancia de este concepto, Kitcher lo retoma en su última obra dándole mayor peso a los valores sociales a lo largo de todo el proceso.

Una investigación puede ser significativa a nivel epistémico y claramente perjudicial a nivel práctico. Reconocer que los intereses sociales, éticos, políticos, etc. juegan un papel determinante

en la actividad científica e introducirlos en el concepto de significación es un paso fundamental para poder ofrecer una descripción naturalizada del conocimiento. El concepto de significación científica es en sí mismo un criterio normativo y valorativo que nos permite evaluar qué prácticas científicas son relevantes para la consecución del bien común. Un criterio que integra elementos políticos, sociales y epistémicos, y que sin minar la racionalidad científica atiende a la pregunta de para qué la ciencia y nos permite desarrollar un modelo de ciencia bien ordenada en el marco de una sociedad democrática.

REFERENCIAS

Goldman, A. 1999, *Knowledge in a Social World*, New York, Oxford University Press.

Kitcher, P. 2001, *Science, Truth and Democracy*, New York Oxford University Press.

Kitcher, P. 2002, 'On the Explantory Role of Correspondence Truth', *Philosophy an Phenomenological Research*, vol. 64, nº 2, 346-364.

Kitcher, P. 2003, 'What Kinds of Science Should Be Done', in Alan Lightman, Dan Sarewitz, and Christina Dresser (eds), *Living With the Genie*, Washington D.C., Island Press, 201-224.

Longino, H. 1990, *Science as Social Knowledge. Values and Objectivity in Scientific Inquiry*, Princeton, Princeton University Press.

Mastroleo, I. 2011, 'La evaluación de la investiagación científica en las sociedades democráticas: Kitcher, Rawls y el enfoque de las verdades significativas', *Revista Redbioética/UNESCO*, año 2, 2(4), 43-60.

LA LÓGICA DE LAS IMÁGENES: RETÓRICA Y ARGUMENTACIÓN EN LAS PRÁCTICAS CIENTÍFICAS REPRESENTACIONALES

EDUARDO ZUBIA
University of the Basque Country UPV/EHU
eduzubia@gmail.com

ABSTRACT: Our aim is to identify good practices in the use of scientific images, resorting primarily to articles recently appeared in *The Astrophysical Journal*. We claim that the explicit description of the methodology employed in the elaboration of a published image will not suffice, and that it should be accompanied by an assessment of the epistemic role played by the image in the argumentation, or in the rhetoric, of the article of which it is part. We propose to identify four distinct epistemic fields involving the production of scientific images: (1) phenomenological images, close to the observational practices, (2) heuristic images, associated with the statistical analysis of empirical data, (3) modelled images, coming from numerical simulations, and (4) analytic images, derived from the theoretical realm. The clear identification of the epistemic scope for each image type facilitates a dialogue between different scientific cultures around a common vocabulary. We think that this epistemological analysis can be extended to other natural sciences.
KEYWORDS: Epistemology, Images, Astrophysics.

Introducción – el papel de la imagen científica

En este trabajo nos interesamos por las imágenes que aparecen en las publicaciones científicas, en particular en el área de la astrofísica. Nuestro objetivo consistirá en identificar buenas prácticas en el uso de las representaciones visuales. La validación de la metodología seguida en la obtención de una imagen forma parte, evidentemente, de las buenas prácticas científicas. Proponemos que también será necesario validar el uso epistémico que se hace de la imagen, explicitando el papel que ésta juega en la argumentación seguida en la publicación. En nuestro estudio emplearemos principalmente artículos de investigación aparecidos recientemente en revistas de referencia del ámbito de las ciencias físicas, como *The Astrophysical Journal*.

1. Epistemología – perspectivas deflacionarias e históricas

Los artículos de Mauricio Suárez de 2003 y 2004 parecen imponerse, al menos en parte de la literatura epistemológica actual, como el punto de partida de un consenso creado en torno a una concepción deflacionaria de la representación científica. Por ejemplo, para Bas van Fraassen tanto

los modelos teóricos como los modelos de datos son ambos entidades abstractas, y de la relación de los modelos de datos con los fenómenos que representan tan sólo podremos decir que es selectiva, y relevante exclusivamente para los fines del investigador (van Fraassen 2008, 253). Tendremos así que el conocimiento científico podrá ser objetivo exclusivamente en la medida en que implique una máxima intersubjetividad (van Fraassen 2008, 266). En este contexto, los instrumentos de observación no serían concebidos como ventanas miméticas al mundo invisible, sino como creadores de nuevos fenómenos observables (van Fraassen 2008, 100). Esta caracterización de la observación empírica necesariamente constriñe el alcance epistemológico de las representaciones visuales.

En la década precedente, Ronald Giere señalaba que tanto para el positivismo lógico como para el relativismo sociológico la imagen científica carece de cualquier valor representacional. Giere propone sin embargo construir una tercera vía epistemológica, examinando precisamente las prácticas visuales de teóricos y experimentadores, más allá por lo tanto del mero análisis proposicional de la tradición analítica (Giere 1996). Más recientemente, Giere ha incidido sobre el papel epistémico de las imágenes astronómicas obtenidas mediante instrumentos de observación contemporáneos. Según Giere, dichas imágenes serían modelos de los datos, derivados de la correcta aplicación de una serie de prácticas científicas, siguiendo un procedimiento que considera interno con respecto a la perspectiva específica del instrumento empleado (Giere 2006, 48-49). En este sentido, lo más concluyente que podrá decirse acerca de un modelo científico es que, dadas las perspectivas observacionales y teóricas asumidas, se ajusta adecuadamente al objeto de estudio (Giere 2006, 92). Este marco teórico general, así definido por van Fraassen y Giere, será el que discutiremos en este trabajo.

La metodología que emplearemos podría caracterizarse con el título de un artículo reciente de Philip Kitcher: "la epistemología está ciega sin la historia" (Kitcher 2011). Según Kitcher, la filosofía de la ciencia debe abandonar la búsqueda de principios a priori a cambio de la verificación de sus propuestas metodológicas mediante el contraste con la historia del conocimiento. En relación con esta línea de trabajo, y con el fin de precisar las herramientas conceptuales que emplearemos, recordamos las siguientes conclusiones epistemológicas de Peter Galison, quien en su estudio sobre la historia de la física de partículas pone de manifiesto la independencia parcial del trabajo de físicos teóricos y físicos experimentales: unos y otros pueden llegar a conclusiones independientes, avaladas por sus propias prácticas, aún a pesar de la posible incompatibilidad de sus producciones respectivas. La forma en que ambas tradiciones acaban consolidando resultados no pasa, enfatiza Galison, por el uso de enunciados de protocolo (positivistas) o el choque de paradigmas

inconmensurables (relativistas), sino por la coordinación local, que requerirá la creación de un idioma epistémico común (Galison 1997, 802-803). Pasemos ahora a analizar algunos usos de la imagen astronómica contemporánea.

2. Discusión – prueba empírica e imagen mimética

Puede ser interesante comenzar analizando una polémica científica: Clowe y colaboradores publicaron lo que denominaron "una demostración empírica directa de la existencia de materia oscura" (Clowe 2006). El argumento central del artículo se formulaba en torno a una imagen de la distribución estimada de materia en un cúmulo de galaxias, formado por dos subcúmulos en colisión. Según esta imagen, el pico de concentración de materia oscura coincide con la distribución visible de galaxias. Sin embargo, el año siguiente Mahdavi y colaboradores, siguiendo técnicas idénticas a las de Clowe y aplicándolas a otro cúmulo, encontraban "un pico de masa sin galaxias que no puede ser fácilmente explicado con el paradigma actual de materia oscura" (Mahdavi 2007).

Señalemos que ambos equipos han seguido una argumentación que puede ser descrita en los mismos términos que emplean van Fraassen y Giere, en tanto que obtienen sus conclusiones de la comparación de modelos de datos (la distribución espacial de materia) con modelos teóricos (simulaciones numéricas de la colisión de los cúmulos). En efecto, la coincidencia de las representaciones derivadas de las observaciones con las simulaciones numéricas confortó a los autores del primer artículo en su conclusión, mientras que, significativamente, el problemático cúmulo Abell 520 ni siquiera ha podido ser modelizado con éxito hasta la fecha (Jee 2012). A pesar de que el primer equipo sigue una metodología científica impecable, su afirmación acerca de la "demostración empírica directa" parece a todas luces excesiva, especialmente teniendo en cuenta el contraejemplo proporcionado por el segundo equipo. Puede pensarse que el carácter aparentemente mimético de las imágenes presentadas ha llevado a los autores a identificar modelos calculados con fenómenos observados, al menos en lo que concierne a la retórica asertiva del título del artículo; quizás un título más adecuado hubiera sido "nuevas observaciones compatibles con los modelos de materia oscura fría".

Llama la atención sin embargo que tanto autores como evaluadores y editores de la principal revista de investigación en astrofísica, personas que por lo tanto definen las buenas prácticas de la comunidad de investigadores, no tengan ningún reparo en asumir el carácter de prueba empírica que puede revestir una imagen, en contraste con la visión deflacionaria de van Fraassen (el instrumento como creador de fenómenos) y Giere (la imagen como modelo de datos). En la siguiente discusión sobre la utilización de imágenes en las ciencias físicas, intentaremos identificar por una parte si las

buenas prácticas científicas nos permiten hablar de algo más que de modelos, y por otra si la epistemología nos puede ayudar a distinguir entre una argumentación legítima y un exceso retórico.

3. Buenas prácticas – actividades científicas representacionales

Proponemos identificar cuatro campos epistémicos diferenciados en cuanto a la producción de imágenes científicas. Señalaríamos (1) las imágenes fenomenológicas, próximas a las prácticas observacionales, (2) las imágenes heurísticas, asociadas al análisis estadístico de datos empíricos, (3) las imágenes modelizadas, procedentes de simulaciones numéricas y (4) las imágenes analíticas, derivadas del campo teórico.

4. Campo fenomenológico

En las imágenes de cúmulos de galaxias obtenidas en los últimos 20 años por el Telescopio Espacial Hubble queda de manifiesto la ubicuidad del fenómeno conocido como lente gravitacional. En contra de las precauciones de van Fraassen y Giere, no hay un solo astrónomo que no considere que estas imágenes representan de hecho un fenómeno real. Según las más sólidas prácticas profesionales de los astrónomos, ciertos objetos observacionales son identificados directamente con fenómenos naturales, no con fenómenos instrumentales o modelos de datos. Los conceptos deflacionarios, aunque puedan ser correctos desde un estricto punto de vista epistemológico, resultan ser en este caso engorrosos. Galison identifica en la física experimental de comienzos del siglo XX una tradición mimética, heredera de la ciencia romántica de Howard y Goethe, y que mediante ingeniosos dispositivos instrumentales pone en relación los fenómenos del mundo natural con las observaciones del laboratorio. La técnica experimental de la cámara de burbujas, que relaciona las trazas de ionización con las trayectorias de las partículas subatómicas invisibles, procede de esta tradición (Galison 1997, 140). Designamos como *fenomenológicas* las imágenes derivadas de esta familia de prácticas. Resaltemos que no se trata de una actividad científica obsoleta: el estudio morfológico visual sigue siendo imprescindible en otros campos ajenos a la física de partículas, como la taxonomía, la astronomía o la dinámica de fluidos. En las publicaciones de estas áreas se utilizan tanto representaciones naturalistas como esquemas sintéticos. Ejemplos característicos de esta categoría llenan las páginas del *Atlas internacional de Nubes*, en forma de fotografías representativas y claves de clasificación (Organización Meteorológica Mundial 1975 1987). El objetivo de esta publicación es establecer criterios para la descripción sistemática del tiempo atmosférico, que es empleada en contextos pragmáticos como los partes para la navegación. Insistimos en el carácter empírico de esta publicación, alejada de cualquier formalismo matemático.

Es más, en áreas más teóricas como la simulación climática las descripciones de este tipo carecen actualmente de utilidad (Zelinka 2012).

5. Campo heurístico

Galison distingue dos tradiciones experimentales fundamentales en la física de partículas. Una de ellas se identifica con la cultura de la imagen, representada por el estudio de las trazas de partículas en las cámaras de niebla, y la otra se identifica con la cultura lógica, caracterizada por el uso de circuitos lógicos en combinación con dispositivos contadores de partículas. A pesar de sus diferencias, ambas tradiciones persiguen objetivos comunes: reconstruir las trayectorias de las partículas que atraviesan los detectores, y acumular evidencias estadísticamente significativas sobre sus propiedades. Durante los años 1960, la automatización del proceso de lectura de los registros desplazó el estudio de las trazas desde el mundo de la imagen hacia el análisis cuantitativo de los instrumentos electrónicos (Galison 1997, 401). Para esta nueva cultura híbrida, el análisis de datos dejó de ser una parte auxiliar del proceso experimental: "el análisis de datos es el experimento" (Galison 1997, 429). En astronomía, retomando nuestra discusión sobre los cúmulos de galaxias, el análisis estadístico de las lentes gravitacionales permite estimar, mediante métodos iterativos de cálculo numérico, un mapa de la posible distribución de materia oscura (invisible) en el cúmulo. En este caso, no puede hablarse de representación directa, y sí está plenamente justificado hablar de modelos de datos. Para el análisis numérico puede utilizarse como dato primario la información morfológica de las imágenes fenomenológicas, pero es habitual también el empleo de imágenes ricas en información cuantitativa, como espectrogramas o figuras de difracción. A su vez, para estas últimas imágenes sí tiene pleno sentido hablar de fenómenos originados en el dispositivo experimental. Entendemos que el rasgo definitorio de este campo epistémico es el análisis estadístico. En efecto, la relación objetiva que pueda establecerse entre estadísticas numéricas y fenómenos observables proporcionará el fundamento de esta práctica, no el análisis de casos individuales significativos. La progresiva adopción de esta familia de técnicas experimentales ha sido definitoria de la ciencia del siglo XX. También es significativa en este campo la utilización de representaciones visuales como resultado del proceso de análisis: esquemas que sintetizan los resultados numéricos y estadísticos, e incluyen habitualmente información morfológica. Atribuimos a estas imágenes la denominación de *heurísticas*. Por ejemplo, puede verse cómo el empleo de las figuras de difracción de rayos X (Friedrich 1912) desplazo las descripciones cualitativas de los atlas mineralógicos de la época romántica en beneficio de la cristalografía, centrada en el estudio de las

relaciones de simetría reveladas por la difracción, y que se caracteriza por el uso de estilizadas representaciones visuales de las retículas cristalinas (Niggli 1920). Recalcamos que en estas prácticas no tiene por qué darse necesariamente una justificación teórica deductiva: la clasificación cristalográfica quedó asentada años antes de que existiera una teoría cuántica de los orbitales moleculares para explicar el origen de las simetrías.

6. Campo simulacional

Como hemos comentado anteriormente, en los artículos de Clowe y Mahdavi la carga de la demostración recae sobre la comparación entre los modelos de datos de distribución de materia oscura y las imágenes generadas en las simulaciones numéricas de colisiones de cúmulos galácticos. Recientemente, Eric Winsberg ha analizado el estatuto epistemológico de las simulaciones numéricas. Winsberg (al igual que Galison) destaca la independencia parcial de las prácticas computacionales, y describe la importancia que éstas han adquirido desde mediados del siglo XX: las simulaciones son una parte necesaria de la metodología científica actual en el estudio de sistemas para los que se carece de soluciones matemáticas analíticas. En cuanto a las imágenes generadas por ordenador, Winsberg menciona el carácter mimético de las representaciones visuales producidas en el estudio de la dinámica de fluidos. Señala que el uso de imágenes realistas en las representaciones de los resultados de las simulaciones es un paso imprescindible para su comparación con los sistemas simulados, como parte del proceso de formulación de inferencias y validación de resultados propio de las prácticas numéricas (Winsberg 2010, 34). Aquí las llamaremos imágenes *modelizadas*. En astrofísica es característico el uso de imágenes miméticas en los modelos numéricos, por ejemplo en las simulaciones de las inestabilidades hidrodinámicas previas a las explosiones de supernova (Endeve 2012). El fenómeno así representado ocurre en el interior de una estrella al borde de la explosión y jamás podrá ser observado directamente. Se estudian sin embargo las similitudes de estas imágenes con las observaciones de restos de supernova (López 2011), y con los modelos materiales empleados en los experimentos de laboratorio de dinámica de fluidos (Foglizzo 2012), desvelándose así líneas de investigación prometedoras.

7. Campo analítico

James Brown ha propuesto en los últimos años un nuevo punto de vista sobre las imágenes matemáticas, oponiéndose a la tradición "iconoclasta" defendida abiertamente por la escuela bourbakiana. Según Brown, algunas de las imágenes empleadas por los matemáticos no pueden entenderse como representaciones de los conceptos teóricos, sino como instrumentos, herramientas

de exploración de las ideas matemáticas (Brown 2008, 40). No entraremos aquí a discutir la defensa del platonismo matemático que hace Brown. Nos interesará sin embargo el uso de la imagen matemática como instrumento de exploración de los conceptos analíticos empleados en las ciencias naturales. Por ejemplo, en el estudio de las colisiones de cúmulos de galaxias, se recurre a esquemas gráficos para explicar los diferentes resultados analíticos obtenidos al utilizar materia oscura de naturaleza fermiónica o bosónica (Lee 2008). Entendemos que estos esquemas no pueden describirse satisfactoriamente como ilustraciones prescindibles de los conceptos teóricos, sino que constituyen herramientas de exploración del pensamiento visual. De ahí nuestra denominación de imágenes *analíticas*. El ejemplo característico de esta categoría serían los diagramas de Feynman, herramienta gráfica ampliamente utilizada para el cálculo de perturbaciones en teoría cuántica de campos. Hay que observar que, a pesar del aspecto de los diagramas, estos no son representaciones de las trayectorias de las partículas subatómicas, sino instrumentos analíticos respaldados por las aproximaciones más tradicionales del cálculo de perturbaciones.

Conclusión – imagen y ciencias naturales

Aunque pensamos que este marco epistemológico es especialmente interesante para la astronomía, también creemos que el análisis puede extenderse a otras ciencias naturales. Por ejemplo, (1) el estudio morfológico del porte arbóreo (Hallé 1968), (2) el análisis estadístico de la distancia genética entre especies (León-Enríquez 2008), (3) la simulación numérica de los esquemas de desarrollo vegetal (Prusinkiewicz 2000) y (4) la teoría botánica que analiza al árbol en términos de unidades de replicación (Barthélémy 1997), proporcionan una visión coherente y fundamentan un esquema de clasificación para la arquitectura arbórea.

En el conjunto de estas actividades, el uso de imágenes facilita la comunicación entre las distintas tradiciones profesionales, y permite establecer un diálogo interdisciplinar entre subculturas científicas en torno a un vocabulario común. En todas estas líneas de trabajo, las representaciones visuales constituyen el vehículo preferente en el que se sintetizan los resultados de la investigación, y son piezas necesarias de la argumentación seguida en las publicaciones. Por lo tanto, la clara identificación del alcance epistémico de cada imagen formará parte de las buenas prácticas científicas.

REFERENCIAS

Barthélémy, D. & al. 1997, ‘Architecture, Gradients Morphogénétiques et Âge Physiologique Chez les Végétaux’, in Bouchon, J. & al. (eds.), *Modélisation et Simulation de l'Architecture des Plantes*, Paris, INRA Éditions, 89-136

Brown, J. 2008, *Philosophy of Mathematics*, Nueva York, Routledge

Clowe, D. & al. 2006, 'A Direct Empirical Proof of the Existence of Dark Matter', *The Astrophysical Journal Letters* 648(2), L109-L113

Endeve, E. & al. 2012, 'Turbulent Magnetic Field Amplification from Spiral SASI Modes', *The Astrophysical Journal* 751(1), id. 26

Foglizzo, T. & al. 2012, 'A Shallow Water Analogue of the Standing Accretion Shock Instability', *Physical Review Letters* 108(5), id. 051103

Friedrich, W. & al. 1912, 'Interferenz-Erscheinungen bei Röntgenstrahlen', *Sitzungsberichte der Königlich Bayerische Akademie der Wissenschaften*, 303-322

Galison, P. 1997, *Image and Logic*, Chicago, The University of Chicago Press

Giere, R. 1996, 'Visual Models and Scientific Judgment', in Baigrie, B. (ed.), *Picturing Knowledge*, Toronto, University of Toronto Press, 269-302

Giere, R. 2006, *Scientific Perspectivism*, Chicago, The University of Chicago Press

Hallé, F. 1968, 'Recherches sur l'Architecture et la Dynamique de Croissance des Arbres Tropicaux', *Sixième Conférence Biennale de la West African Science Association*, 12 p.

Jee, M. & al. 2012, 'A Study of the Dark Core in A520 with Hubble Space Telescope', *The Astrophysical Journal* 747(2), id. 96

Kitcher, P. 2011, 'Epistemology Without History is Blind', *Erkenntnis* 75(3), 505-524

Lee, J. & al. 2008, 'BEC Dark Matter Can Explain Collisions of Galaxy Clusters', *arXiv.org* 0805.3827v1 (documento electrónico)

León-Enríquez, B. & al. 2008, 'The Architecture of Phyllanthus Acuminatus Vahl', *Adansonia* 30(1), 137-149

López, L. & al. 2011, 'Using the X-ray Morphologies of Young Supernova Remnants to Constrain Explosion Type, Ejecta Distribution, and Chemical Mixing', *The Astrophysical Journal* 732(2), id. 114

Mahdavi, A. & al. 2007, 'A Dark Core in Abell 520', *The Astrophysical Journal* 668(2), 806-814

Niggli, P. 1920, *Lehrbuch der Mineralogie*, Berlín, Gebrüder Borntraeger

Organización Meteorológica Mundial 1975, *International Cloud Atlas - Volume I*, Ginebra, Organización Meteorológica Mundial

Organización Meteorológica Mundial 1987, *International Cloud Atlas - Volume II*, Ginebra, Organización Meteorológica Mundial

Prusinkiewicz, P. & Remphrey, W. 2000, 'Characterization of Architectural Tree Models Using L-Systems and Petri Nets', *4th International Symposium on the Tree*, 177−186

van Fraassen, B. 2008, *Scientific Representation*, Oxford, Oxford University Press

Winsberg, E. 2010, *Science in the Age of Computer Simulation*, Chicago, The University of Chicago Press

Zelinka, M. & al. 2012, 'Computing and Partitioning Cloud Feedbacks using Cloud Property Histograms - Part I', *Journal of Climate* (en imprenta)

HISTORY OF SCIENCE

HISTORIA DE LA CIENCIA

PANOFSKY: EL CONFLICTO ENTRE LA PERSPECTIVA LINEAL Y LA PERSPECTIVA ANGULAR

CARLOS ALBERTO CARDONA
Escuela de Ciencias Humanas
Universidad del Rosario
Bogotá-Colombia,
carlos.cardona@urosario.edu.co.

RESUMEN: Se defiende que el conflicto que anuncia Panofsky entre la proposición 8 de la Óptica de Euclides y la Perspectiva Lineal carece de fundamento. Se mostrará que las reglas de construcción perspectiva sugeridas por Piero della Francesca usan las definiciones de Euclides y armonizan sin dificultad la proyección perspectiva y la proposición 8. En consecuencia si el surgimiento del espacio moderno exige una ruptura con el espacio psicofisiológico, como lo sugiere Panofsky, esta ruptura se debe defender con argumentos distintos al de esgrimir el conflicto aludido.

En la soberbia obra de Panofsky *La perspectiva como forma simbólica*, el autor se propone mostrar que los principios de la óptica de Euclides resumen las propiedades del espacio psicofisiológico, mientras que los principios de la perspectiva lineal, desarrollada en el Renacimiento, ofrecen un eslabón intermedio entre el espacio psicofisiológico y el espacio matemático propio de la ciencia moderna. Panofsky sugiere que los principios de la perspectiva lineal no sintetizan los principios o leyes de la percepción tal y como ella se da. Muy a nuestro pesar, tales principios (los de la perspectiva lineal) sintetizan formas convencionales de aprehender u organizar nuestra experiencia perceptual.

La perspectiva lineal, cree el autor, debe erigirse desmontando la proposición 8 de la óptica de Euclides. Me propongo mostrar que la perspectiva lineal no se encuentra en conflicto con la proposición 8 y que teóricos del Renacimiento como Piero della Francesca lograron incorporar sin conflicto alguno las demandas de Euclides y las reglas de proyección pictórica. En la primera parte

presento el argumento de Panofsky; en la segunda parte ofrezco el argumento de la construcción perspectiva de Piero della Francesca y muestro que no existe el conflicto señalado por Panofsky.

1. El argumento de Panofsky

El argumento se puede desplagar en cuatro fases.

Primera fase: caracterización de un espacio racional. Panofsky caracteriza el *Espacio Racional* como infinito, constante y homogéneo. En el espacio matemático no hemos de reconocer limitación alguna. En ese sentido, dicha estructura debe aparecernos como una estructura sin límites (infinita). Ahora bien, la percepción natural, cree Panofsky, desconoce tal carencia de límites. De otra parte, que el espacio matemático sea homogéneo significa que sus elementos, a saber los puntos individuales, tan sólo señalan posiciones potenciales. Ningún punto se diferencia de otro por razones sustanciales; se diferencia sólo en virtud de las relaciones de lugar que guarda con otros puntos. El espacio de la percepción natural carece de la pretendida homogeneidad, toda vez que las diferenciaciones arriba-abajo, derecha-izquierda, adelante-atrás imponen la marca característica de nuestras aprehensiones visuales y táctiles. Finalmente, dado que los puntos carecen de toda determinación, hemos de esperar que no se de en ellos una especie de devenir que modifique propiedad sustancial alguna. Dado que el contenido perceptual asociado con las nociones de arriba-abajo, derecha-izquierda, delante-detrás, cerca-lejos cambia en la medida en que vemos con dos ojos fijos en una cabeza que se desplaza y gira provocando con ello variaciones cinestésicas que covarían con las modificaciones en el campo visual, hemos de advertir que no hay nada constante en el espacio psicofisiológico.

Segunda fase: construcción de la geometría del espacio psicofisiológico. ¿Existe o se puede construir una geometría para el espacio psicofisiológico? El núcleo central de la respuesta y la argumentación de Panofsky depende de la interpretación que él le da a la proposición 8 de la Óptica de Euclides. El autor asume que dicha proposición sistematiza un presupuesto básico que la Antigüedad mantuvo firmemente, a saber:

> 'las dimensiones visuales (en tanto proyecciones de las cosas sobre la esfera ocular) no están determinadas por la distancia existente entre los objetos y el ojo, sino exclusivamente por la medida del ángulo visual (de ahí que sus relaciones sean expresadas sólo mediante grados angulares, medidos con exactitud, o mediante arcos de círculo, y no mediante simples medidas lineales)' (p. 19).

Panofsky sostiene que la proposición 8 de Euclides se opone radicalmente a los principios de la perspectiva lineal y respeta en forma más fina la geometría propia del espacio psicofisiológico.

Recordemos, entonces, la proposición 8 de Euclides: «*Las magnitudes iguales y paralelas situadas a distintas distancias del ojo no se ven proporcionalmente a las distancias*» (2000, prop. 8). Veamos ahora en doble columna los presupuestos de la perspectiva lineal y los de perspectiva angular que parece ajustarse a las demandas de la proposición 8 de Euclides (Figura 1). Sean *BC* y *DE* dos segmentos congruentes que pueden disponerse a distancias diferentes *AB* y *AD* de un observador *A*. Bajo las expectativas de la perspectiva lineal, las longitudes de las imágenes visuales se contemplan en el corte de un plano con la pirámide visual. Así entonces, las longitudes *BC* y *DE* se contemplan de acuerdo a las longitudes *KS*, la primera, y *SJ*, la segunda. Bajo las expectativas de la perspectiva del espacio psicofisológico, dichos segmentos se comparan a partir de las diferencias angulares α y $\alpha+\beta$.

Perspectiva lineal

$BC \cong DE$; $AB =_{Def} d_1$; $AD =_{Def} d_2$

$$\frac{KS}{SJ} = \frac{d_2}{d_1}$$

Perspectiva angular

Apariencia de *BC*: α; Apariencia de *DE*: α+β

$$\frac{\alpha}{\alpha+\beta} > \frac{d_2}{d_1}$$

Figura 1. Comparación de las perspectivas lineal y angular

Tercera fase (Reconocer prácticas artísticas identificadas con la perspectiva propia del espacio psicofisiológico). Panofsky pregunta si la Antigüedad Clásica contaba con prácticas artísticas que se mantuvieran fieles a la proposición 8 de Euclides. La respuesta de Panofsky es afirmativa. El autor interpreta así la recomendación de Vitrubio que consiste en curvar algunos elementos arquitectónicos para compensar aberraciones marginales (p. 88, n. 11). Panofsky sugiere también interpretar el famoso pasaje de Vitrubio como si estuviese recomendando una proyección esférica.

Cuarta fase (Señalar los elementos de transición que van desde el espacio psicofisológico al espacio matemático). Panofsky estudia los que, a su juicio, son los pasos que permiten la transición

desde la perspectiva angular hasta la perspectiva lineal. El autor se detiene en varias formas de expresión artística propias del Medioevo, analiza qué se conserva de la Antigüedad Clásica y qué se modifica en sus matices.

2. La construcción perspectiva de Piero della Francesca y la deconstrucción del aparente conflicto.

Pretendo mostrar que Panofsky se equivoca cuando aduce que los teóricos del Renacimiento suprimen por completo el Teorema 8 o lo enmiendan parcialmente. Afirma el autor: «*Probablemente es por algo y no por pura casualidad que más tarde el Renacimiento, cuando parafrasea a Euclides* […]*, bien suprime totalmente este octavo teorema, bien lo "enmienda" parcialmente, de modo tal que le hace perder todo su sentido originario*» (p. 19). Mostraré que en el tratado de Piero della Francesca *De prospectiva pingendi* se ofrecen reglas de representación pictórica que se defienden a partir de los axiomas de Euclides y armonizan con los criterios de la perspectiva lineal.

2.1 El argumento de Piero

El problema original que le interesa a Piero puede formularse en los siguientes términos: dados un cuadrado tendido en el plano de base, un velo pictórico perpendicular a dicho plano y, en una ubicación que se encuentra claramente determinada, un observador al otro lado del cuadrado con respecto al velo, se pide hallar el trapecio que se obtiene como resultado de degradar el cuadrado en el velo pictórico. Sea, entonces, el cuadrado *BCFG*, con *BC* sobre la línea de base, el velo pictórico perpendicular al plano de base por la recta *BC* y un observador cuya ubicación se establece a partir de la tripleta (*d*, *m*, *h*) se pide determinar las características geométricas del trapecio *BCF´G´* (Figura 2). Determinar las características del trapecio significa hallar: (i) la magnitud del segmento *F´G´*, *x*, (ii) la magnitud de la altura *F´F´´*, *y*, (iii) la magnitud de *BF´´*, *z*.[1]

[1] Por razones de espacio, me abstengo de presentar el cálculo de z. Esto no representa una limitación en el argumento que esgrimiré.

Figura 2. Formulación del problema

La construcción de Piero consta de dos partes que después se integran en un único esquema. La primera parte determina las dimensiones que debe tener el segmento superior *x*. La segunda parte determina la altura del trapecio *y*.

En la primera parte se pide que imaginemos, por ahora, que la altura del observador se reduce hasta coincidir con el punto *M*.[2] Se puede ver que, sin importar cuál es la magnitud del segmento *DM*, en todos los casos la longitud de *F´´G´´* es la misma y depende únicamente de la longitud de los lados del cuadrado y de la distancia *DB*. La demostración, que Piero omite, es muy simple. Sean *BC, CG, GF, FB* los lados de un cuadrado, cada uno de ellos de longitud *L* (Figura 3). Los triángulos *MGF* y *MG´´F´´* son semejantes y por lo tanto:

$$(1) \quad \frac{x}{L} = \frac{d}{L+d}$$

[2] En esta posición, todos los objetos en al plano de base degradan en segmentos sobre la línea de base.

Figura 3. Degradación de GF

Del resultado se infiere que la longitud *x* es independiente de *m*.[3] Piero aduce que desde el punto de vista de *M*, *F´´G´´* se contemplará como si aparentara el mismo tamaño que *FG* toda vez que el ángulo de contemplación es el mismo. Piero se adecúa al principio central de Euclides que sugiere que el tamaño aparente de los objetos contemplados depende de la amplitud angular de la pirámide visual. Este predicamento euclidiano se presenta como la proposición II del libro I del *De Prospectiva*. Transcribo a continuación la explicación de Piero ajustando la nomenclatura a la que estoy empleando:

> «Y la misma proporción de [F´´G´´] a [FG] es [dada por] [MF´´] a [MF]…[4] Y cuando las distancias y las cosas están en la misma proporción como la altura del ojo y la cosa tal y como es degradada, es claro que la degradación es correcta» (Prospectiva, I, Prop. XIII).

Esto resuelve el primer componente de la tripleta buscada.

En la segunda parte se establece la altura del trapecio. Si ahora, sin cambiar la altura del observador, movemos *M* hasta hacerlo coincidir con *D*, el corte *F´´* coincide ahora con *B*. La recta *AF* corta al velo pictórico en un punto F_o de tal manera que BF_o determina la altura del trapecio (Figura 4). Para defender la propuesta, Piero acude a dos argumentos inspirados en la óptica de Euclides: (i) si dos rectas paralelas (en este caso *BC* y *FG*) son perpendiculares a una tercera (en este caso *DF*) y a partir de *D* se traza una perpendicular al plano de las otras rectas para establecer sobre ella la posición de un observador (en esta caso *A*), ocurre que desde *A* se percibe al segmento *FG* (el más alejado) por encima de la percepción de *BC* (*De prospectiva,* I, prop. VI).[5] (ii) *BF*, visto

[3] También se puede probar que el resultado es independiente de la altura del observador.
[4] Es claro que *MF´´* y *MF* están en la misma proporción que *DB* y *DF*.
[5] Cfr. Euclides (2000), prop. 10.

desde *A*, tiene la misma apariencia que BF_o porque ambos se contemplan bajo el mismo ángulo (*De prospectiva*, I, prop. II).[6] Sea $BF_0 = y$, de la semejanza de los triángulos *ADF* y F_0BF se impone:

$$(1)\ \frac{h}{y} = \frac{L+d}{L}$$

Si las transversales degradan en transversales (asunto que no demuestra Piero), la altura del trapecio es independiente de la magnitud *m*. Esto resuelve el segundo componente de la tripleta buscada.

Figura 4. Altura del trapecio

Piero reúne después las dos construcciones anteriores en una sola figura y sugiere su método particular para degradar un cuadrado (Figura 5). Presento a continuación la síntesis respetando la nomenclatura que he propuesto. El lector debe estar atento al cambio de papel que desempeña la hoja sobre la que se hace el dibujo: en cada caso representa un plano diferente.

El papel visto como un plano perpendicular tanto al plano de base como al velo pictórico. Piero asume que la recta *DBF* es una ortogonal al velo pictórico, con *DB=d*, y el segmento perpendicular *DA* representa la altura del observador, *h*. Hallamos ahora el corte de *AF* con la perpendicular por *B*, sea este corte el punto F_0. BF_0 define la altura del trapecio que ha de representar el cuadrado degradado.

[6] Cfr. Euclides (2000), definición 4.

El papel visto como velo pictórico. Dada la nueva función del papel, rebautizamos el punto *F* como punto *C* y asumimos *BC* como línea de base. Trazamos ahora una paralela a *BF* por F_0, y fijamos así la altura a la que debe aparecer el segmento *F´G´*.

El papel visto como plano de base. Imaginamos el cuadrado en sus dimensiones originales *BFGC* (*BC* la línea de base).[7] A partir de *A* trazamos las rectas *AF* y *AG* para determinar los cortes con *BC*, a saber: *F´´*, *G´´*. La longitud del segmento *F´´G´´* determina la longitud del segmento superior del trapecio que estamos buscando.

El papel nuevamente visto como velo pictórico. Sobre la paralela por F_0 a la línea de base, que demanda rebautizar nuevamente el punto *F* como punto *C*,[8] escogemos un punto *F´* y a partir de él construimos el segmento *F´G´* de idéntica longitud que *F´´G´´*.[9] El trapecio que vamos a construir por encima de esta línea muestra la representación pictórica en el velo de proyección. El trapecio *BF´G´C* es la presentación de un cuadrado degradado para un observador que se encuentra a una distancia *d* del velo y tiene una altura *h*.

Figura 5. Síntesis de Piero

Piero muestra que si se prolongan *BF´* y *FG´* ellos se cortan en un punto *I* que se encuentra a la misma altura del punto *A* (Figura 6).

[7] El punto *F* recupera el papel inicial.

[8] Este ejercicio se facilita toda vez que $BC \cong BF$

[9] La elección de *F´* depende de *z*. El resultado que nos interesa, sin embargo, es independiente de dicha elección.

Figura 6. Convergencia de las ortogonales

2.2 La deconstrucción del argumento de Panofsky

Ahora que sabemos degradar un cuadrado de acuerdo a la regla de proyección de Piero della Francesca, imaginemos dos segmentos congruentes *AB* y *CD* tendidos sobre el plano de base en una distribución transversal (Figura 7). *O* define la ubicación de un observador de altura *OH* ubicado a una distancia *HE* del velo pictórico.[10] Por razones pedagógicas, me permito imaginar que $HA = 2HE$, $HC = 4HE$, esto es, *CD* se encuentra dos veces más lejos del observador que *AB*. Teniendo en cuenta el protocolo de construcción perspectiva de Piero, es claro que *A´B´* y *C´D´* definen las longitudes adecuadas de los segmentos *AB* y *CD* en el velo pictórico. La explicación de lo anterior deviene del hecho de que los segmentos *AB* y *A´B´* se contemplan bajo el mismo ángulo desde el punto *O*; así como ocurre con los segmentos *CD*, *C´D´* (Definición 4 de Euclides). Ahora bien, es claro que $A'B' = 2C'D'$ así como $HA = \frac{HC}{2}$; en otras palabras: las longitudes que han de representar segmentos paralelos de idéntica longitud guardan entre sí la misma proporción que el inverso de las distancias al observador (como exige la perspectiva lineal). También es claro que $\angle B'OA' \neq 2\angle D'OC'$ como demanda la proposición 8 de Euclides. Como vimos en el apartado anterior, en la construcción de Piero se construyen $A''B'' \cong A'B'$ y $C''D'' \cong C'D'$ como los escorzos buscados de los segmentos *AB* y *CD*. La representación en escorzo preserva entonces la exigencia de la perspectiva lineal, a saber $A''B'' = 2C''D''$ ya que representan objetos de idéntica longitud ubicados de tal manera que *CD* se encuentra, con respecto al observador, dos veces más lejos que *AB*. También preserva el hecho de que la apariencia euclidiana de *AB* (ángulo *A´OB´*) no guarda la

[10] Conviene advertir que, visto desde arriba, *OH* no representa la altura del observador, sino la ubicación de los pies del observador con respecto a la ortogonal *EC*. Sin embargo las magnitudes de *A´B´* y *C´D´* son independientes de la longitud *OH*. Así entonces, podemos tomar *OH* con las mismas dimensiones de la altura del observador sin que ello cambie las conclusiones.

proporción inversa con la apariencia euclidiana de CD (ángulo $C'OD'$). Es importante anotar que, si bien en la representación en escorzo ocurre que $A''B'' = 2C''D''$, no es cierto que $EC'' = 2EA''$. Es decir, las dimensiones de la representación de objetos ortogonales no decrece al mismo ritmo que las dimensiones de la representación de objetos transversales.

Figura 7. Escorzo de Piero della Francesca

La confusión de Panofsky se origina por creer que las apariencias ("magnitudes") a las que alude Euclides remiten a la longitud de objetos físicos (imágenes) que se constituyen en el correlato de nuestra percepción: objetos que contemplamos de primera mano en un escenario y que fungen como sustitutos de objetos que existen allende dichos escenarios y, por ello, no pueden ser objeto de contemplación directa. Cuando en realidad Euclides alude, más bien, a la amplitud angular de la pirámide visual. En otras palabras, percibir parece ser, a los ojos de Panofsky, una relación que se establece entre un observador peculiar y una imagen que copia isomórficamente las longitudes de un objeto. Cuando imaginamos un objeto dado, objeto que despliega un proceso de multiplicación de sus especies, nos esforzamos a continuación por estudiar cómo es que finalmente se recogen las especies de dicho objeto, bien sea en la cara posterior del cristalino o en la retina, y llegamos a creer que allí se inicia un segundo proceso de contemplación llevado a cabo por un segundo observador que ahora no es un observador físico y que tiene en su horizonte, ya no al objeto inicial sino a una imagen del primero. Ver, ya no es entrar en contacto directo con un objeto, es entrar en relación con una imagen suya recogida en un telón o escenario. Esta figura está bellamente recogida en el segundo observador que supone Leonardo (Figura 8).

Figura 8. Leonardo, 1968, § 20

REFERENCIAS

EUCLÍDES. (2000). *Óptica.* Madrid: Editorial Gredos.

FRANCESCA, Piero della. (1984). *De Prospetiva Pingendi*. Firenze: Casa Editrice Le Lettere. Edición crítica a cargo de G. Nicco-Fasola.

PANOFSKY, Erwin. (2003). *La perspectiva como forma simbólica.* Barcelona: Tusquets Editores.

VINCI, Leonardo da. (1968). *Manuscript D in The Biblioteque de l´institut de France.* Ver D. S. Strong, *Leonardo da Vinci on the Eye: the MS D in The Biblioteque de l´institut de France.* Disertación doctoral Universidad de California.

APUNTE ACERCA DE LA FILOSOFÍA DE LA CIENCIA DE BLAS CABRERA: EL PAPEL DE DUHEM EN SU DEFENSA DE LA RELATIVIDAD

PABLO RUIZ DE OLANO
University of Notre Dame
USA

ABSTRACT: Blas Cabrera, the so-called father of Spanish physics, devoted many of his efforts during the 1920s to popularizing and defending Einstein's general theory of relativity. In this paper, I make the argument that Cabrera's activities as an introductor of relativity included a philosophical defense of Einstein's theory, which was based on a form of empiricism inspired by Pierre Duhem's holistic conventionalism. I look at Cabrera's 1923 book *Principio de Relatividad* as well as at two conferences given by him in 1921 to conclude that his defense of relativity was conventionalist in the sense that it allowed that different theories may account for empirical data and that choice between them is a matter of convention, but contended that considerations such as simplicity and unity still favor Einstein's theory over those of his critics.
KEYWORDS: Blas Cabrera, relativity, conventionalism, Duhem, Einstein

Introducción

El físico canario Blas Cabrera, a menudo considerado el padre de la física española, fue una figura clave en la recepción de la Relatividad General en España. A pesar de que la figura de Cabrera ha recibido cierta atención en la literatura (Glick 1988; Sánchez Ron y Roca-Rosell 1993), los aspectos filosóficos de su pensamiento han pasado hasta ahora desapercibidos[1]. En este trabajo, defiendo la tesis de que las actividades de Cabrera como introductor de la relatividad en España incluyeron una defensa filosófica de la validez de la teoría de Einstein, que se basó en el convencionalismo holista de Pierre Duhem.

1. El Análisis de Cabrera del Experimento de Eddington en 1919

Entre las actividades de Cabrera como divulgador y defensor de la teoría de Einstein cabe destacar la publicación de su primer libro, *Principio de Relatividad*, que coincidió con la visita de Einstein a España en 1923 (Glick 1988). *Principio de Relatividad* es interesante porque en él,

[1] Glick dedica cierta atención a este aspecto de la obra de Cabrera y de otros autores de la época, aunque su trabajo se centra en el papel de Cabrera como introductor de la relatividad en España. En concreto, el carácter convencionalista del pensamiento de Cabrera no se discute en su libro (Glick 1988).

Cabrera no se limita introducir los fundamentos de la teorías especial y general de la relatividad, sino que ofrece además una defensa filosófica de la teoría de Einstein.

El mejor ejemplo de esto se encuentra tal vez en el capítulo sexto del libro, que está dedicado a teoría de la gravitación de Einstein. Después de derivar la célebre predicción del físico alemán de que los rayos de luz deben desviarse de sus trayectorias rectilíneas en presencia de cuerpos masivos, Cabrera centra su atención en el famoso experimento de Eddington en 1919. El análisis de los resultados de Eddington que proporciona a continuación es interesante porque revela el carácter convencionalista del pensamiento de Cabrera.

Cabrera comienza la discusión mencionando algunas de las objeciones con que se había puesto en entredicho la conclusión de Eddington de que el desplazamiento en la posición de las estrellas cercanas al sol observado durante el eclipse se debía a un efecto relativista. Entre dichas objeciones, Cabrera menciona la escasa magnitud del efecto, de unas pocas décimas de milímetro solamente; la hipótesis de que la deflexión de los rayos de luz responsable del cambio de posición aparente de las estrellas observado durante el eclipse pudiera deberse a una atmósfera solar hasta entonces desconocida; y la propuesta análoga de que pudiera deberse al llamado efecto de refracción anual que había sido descubierto poco antes por Couvoisier.

Cabrera rechaza estas interpretaciones del experimento de Eddington contrarias a la relatividad de la siguiente forma:

> "Dejando a un lado la refutación de cada una de estas hipótesis particulares con que se quiere reemplazar la teoría del físico alemán, he de señalar simplemente su carácter específico para la sola interpretación de este fenómeno, sin que encuentren apoyo en ninguno otro, y menos parentesco ostensible con las interpretaciones, también ad hoc, a que me refería en el caso del movimiento del perihelio de Mercurio y el corrimiento de las rayas espectrales." (Cabrera 1923, 255-56)

Como este fragmento pone de manifiesto, Cabrera no pensaba que tales objeciones a la interpretación relativista de Eddington fueran válidas. ¿Pero en qué argumentos basaba Cabrera su rechazo a estas críticas? En primer lugar, conviene tener en cuenta que la respuesta de Cabrera no consiste en negar la veracidad de las hipótesis propuestas por los críticos de Einstein, dado que su refutación es "dejada a un lado" en el fragmento citado más arriba. Y en cualquier caso, Cabrera aclara inmediatamente que su estrategia no consiste en negar que tales hipótesis puedan ser ciertas:

> "Esto no es negar la posibilidad, y aun la probabilidad, de que en fenómenos tan complejos como el paso de la luz por las vecindades del disco solar y a través de la región de nuestra atmósfera perturbada por el paso del cono de sombra lunar, no intervengan acciones que influyan sobre el resultado perceptible, aunque sólo como términos correctivos." (Cabrera 1923, 257)

Si no en la falsedad de las hipótesis usadas por los rivales de Einstein, ¿en qué basaba Cabrera su defensa de la relatividad? Como el texto citado más arriba pone de relieve, las reservas de Cabrera parecen estar relacionadas con el carácter ad hoc de las alternativas propuestas por ellos. De hecho, en capítulos anteriores, Cabrera analiza las otras dos fuentes de evidencia experimental en favor de la relatividad general disponibles en aquella época: la precesión del perihelio de Mercurio en el capítulo cinco y el corrimiento al rojo de las rayas espectrales de la luz proveniente de estrellas muy masivas en el capítulo seis. En ambos casos Cabrera menciona algunas de las alternativas propuestas por autores escépticos con el trabajo de Einstein, y las rechaza por su carácter artificial y por crear más problemas de los que resuelven.

Por el contrario, Cabrera afirma que las explicaciones proporcionadas por la relatividad general no sólo son mucho más elegantes, sino que también proporcionan un tipo de unidad entre fenómenos que se creían independientes que se echa de menos en las alternativas ofrecidas por sus críticos:

> " Precisamente en la unidad que la teoría einsteiniana forma con los primeros hechos, aparentemente tan dispares, se halla su más fuerte sostén. Ellas son dos consecuencias que de la teoría surgen, sin que en ningún momento de su desarrollo hayan servido de faro que guiase los razonamientos de su autor, ni tampoco sirviesen para dar valor a constantes que en los cálculos se hubiesen conservado indeterminadas." (Cabrera 1923, 256)

Es decir, el argumento decisivo a favor de la relatividad lo proporcionan consideraciones relacionadas con conceptos como unidad y simplicidad, y no el hecho de que las alternativas propuestas por los críticos de Eddington y Einstein sean incompatibles con la observación.

De esta forma, el carácter convencionalista del pensamiento de Cabrera es evidente dado que, al admitir que diferentes interpretaciones del experimento de Eddington son posibles, Cabrera parece admitir que la elección entre ellas es una cuestión de convención. Como hemos visto, su defensa de la relatividad consistió en afirmar que, pese a todo, ciertas convenciones son preferibles en virtud de su mayor simplicidad y elegancia.

2. **La influencia de Duhem en la conferencia de 1921**

El sentido en que el convencionalismo de Cabrera está inspirado en el trabajo de Pierre Duhem es más evidente al compararlo con otras formas de convencionalismo. Henri Poincaré, por ejemplo, había propuesto en su célebre *La Science et l'Hypothèse* que determinados enunciados científicos no pueden ser refutados empíricamente, dado que se trata de definiciones enmascaradas que se adoptan por convención. Pero Poincaré limitó su análisis a determinados enunciados

aparentemente necesarios, como los axiomas de la geometría o los principios de la mecánica clásica (Poincare 1902).

Por el contrario, Duhem afirmaba que, al menos en ciencias experimentalmente densas como la física, el mismo tipo de inmunidad frente a la experiencia afecta a todos los enunciados de una teoría científica y no sólo de un determinado subconjunto de los mismos. Según Duhem, tal conclusión se sigue de un "sencillo análisis lógico" del método científico, de acuerdo con el cual ninguna hipótesis puede ser contrastada con la experiencia de manera aislada, dado que un resultado negativo siempre puede achacarse tanto a la falsedad de la hipótesis a comprobar como a cualquiera de las "hipótesis auxiliares" usadas en la derivación de la predicción en cuestión. Como es bien sabido, Duhem derivó de este razonamiento la conclusión holista de que sólo las teorías científicas en su totalidad, y no hipótesis aisladas, pueden ser contrastadas empíricamente (Duhem 1914).

Esta distinción proporciona un primer indicio del duhemianismo latente en el pensamiento de Cabrera, dado que en los pasajes citados en la sección anterior no se hace ningún esfuerzo por restringir el carácter convencional de las teorías científicas a un determinado tipo de enunciados. Al contrario, la elección en los ejemplos que mencionamos anteriormente era entre diferentes alternativas tomadas en su conjunto.

En cualquier caso, disponemos de evidencia textual que parece confirmar la hipótesis de que el convencionalismo de Cabrera estaba inspirado en la variante holista representada por Duhem. En 1921, Cabrera pronunció una conferencia sobre la teoría de la relatividad en la Sociedad Oceanográfica de Gipuzcoa en San Sebastián. Al igual que en su libro de 1923, Cabrera comenzó su discusión sobre la validez de la evidencia empírica a favor de la relatividad enumerando algunas de las alternativas propuestas por aquellos que eran contrarios a la teoría de Einstein. Pero en 1921, el ascendente duhemiano del pensamiento de Cabrera resulta más evidente. Hablando de los adversarios de Einstein, Cabrera dijo:

> " Primero se trató de negar la exactitud de las observaciones; pero después se ha querido buscar una explicación ad hoc. Esto es sumamente inútil; se puede explicar un fenómeno solo de muy diversas maneras, pero el conjunto es lo que es más difícil de armonizar. Suponer una atmósfera alrededor del Sol, refractora de los rayos en proporción conveniente para que produzca un efecto observado, no cuesta nada; pero si se tiene en cuenta que éste es un hecho que ha de ser explicado, con otros muchos que van por delante, se ve lo inconsistente de tan apriorística manera de concebir." (Cabrera 1921; citado en Glick 1988, 144)

Como puede apreciarse, el argumento de Cabrera en este pasaje es que la dificultad al crear nuevas teorías científicas consiste en armonizar las explicaciones dadas a diferentes fenómenos. Por

un lado, esto sugiere que pueden existir diferentes formas de llevar a cabo tal proceso de "armonización" y que la elección entre las distintas alternativas es una cuestión de convención. Por otra parte, al igual que en *Principio de Relatividad*, Cabrera no hace ningún esfuerzo por restringir dicha elección a un determinado conjunto de principios fundamentales; más bien parece que, para Cabrera, la elección concierne precisamente el cómo se armonizan las diferentes partes de la teoría en cuestión.

3. Cabrera y la Relación entre Geometría y Experiencia

Aunque Einstein constituye un intermediario más que probable en este caso, otro aspecto del pensamiento de Cabrera en el que la influencia de Duhem es notoria es su tratamiento de la relación entre geometría y experiencia. Éste era un aspecto importante de su defensa de la Relatividad General, dado que el tratamiento geométrico dado por la teoría de Einstein a los fenómenos inerciales y gravitatorios exigía una repuesta a la pregunta de cómo es posible que teorías abstractas en geometría puedan usarse para describir el mundo físico.

Cabrera expuso sus puntos de vista al respecto en una segunda conferencia pronunciada en 1921 en la Universidad Central de Madrid. Cabrera empezó por distinguir entre dos tipos de geometría, que llamó "racional" y "experimental". Acerca de esta última dijo:

> " (…) esta geometría, que es la que aplica el agrimensor, el topógrafo, el arquitecto, el constructor de toda especie, no ha de confundirse con aquella ciencia matemática que lleva el mismo nombre y que forma parte de la enseñanza más elemental. Esta otra ciencia eminentemente lógica es, como todas las matemáticas, una construcción de nuestro espíritu." (Cabrera 1921; citado en González de la Posada 1994, 131)

Siendo esto así, ¿cómo es posible que geometrías "racionales" como la descrita por la teoría General de la Relatividad puedan usarse para hacer predicciones sobre el mundo real? ¿Y cómo determinar qué geometría "racional" describe la verdadera geometría del mundo? En principio, parece que las propiedades geométricas del mundo físico podrían medirse de forma directa:

> " Diréis: puesto que se trata aquí de un hecho real que podemos conocer experimentalmente, tendremos modo de escoger entre aquellas geometrías lógicamente posibles, la única que se ajuste a la Naturaleza." (Cabrera 1921; citado en González de la Posada 1994, 133)

Sin embargo, Cabrera rechazó inmediatamente tal posibilidad, arguyendo que:

> " (…) la realización práctica de esta idea tropieza con dificultades que en general rebasan nuestras posibilidades. La medida de aquellas longitudes no puede ejecutarse exactamente. Lleva aparejados errores más o menos grandes que

> engendran una incertidumbre en el valor experimental de la relación que nos ocupa, dentro de cuyos límites puede hallarse la solución del problema." (Cabrera 1921; citado en González de la Posada 1994, 133)

La objeción de Cabrera aquí es que la inevitabilidad del error experimental invalida la estrategia evidente de llevar a cabo mediciones con objecto de descubrir qué geometría de entre las muchas posibles se adecua a la experiencia. ¿Pero cuál es exactamente el problema identificado por Cabrera? ¿Y en qué sentido sugiere esto un nexo con la obra de Duhem?

Para responder a estas preguntas, es importante tener en cuenta que la conferencia de Cabrera en San Sebastián guarda una profunda semejanza con otra conferencia, publicada después bajo el título "Geometría y Experiencia", que el mismo Einstein había pronunciado en la Academia Prusiana de Ciencias tan sólo unos meses atrás. Como Cabrera, Einstein había distinguido entre geometría "práctica" y "puramente axiomática", y se había preguntado cómo es posible que una geometría "puramente axiomática" pueda usarse para formular enunciados acerca del mundo. Einstein también había considerado la posibilidad de determinar empíricamente qué geometría axiomática se aplica al mundo físico y, al igual que el físico canario, había desechado tal posibilidad.

Sin embargo, Einstein reprodujo los detalles de su argumento en detalle, y aludió explícitamente el carácter convencionalista del mismo:

> " (...) los cuerpos sólidos reales presentes en la naturaleza no son rígidos, porque su comportamiento geométrico, es decir, sus posibilidades de disposición relativa, dependen de la temperatura, de fuerzas externas. Así, la relación inmediata entre Geometría y realidad física aparece destruida, y nos vemos forzados a aceptar un punto de vista más general, que caracteriza las postura de Poincaré." (Einstein 1922, 34)[2]

Además, Einstein derivó de este razonamiento la siguiente conclusión de carácter holista:

> La Geometría (G) no predica nada acerca de las relaciones entre objetos reales, sino la Geometría junto con (P) de leyes de la física. Simbólicamente, podemos decir que solamente la suma (G) + (P) está sujeta al control de la experiencia. Así, (G) puede elegirse arbitrariamente, y también partes de (P); estas leyes son convenciones. Todo cuanto hace falta para evitar contradicciones es escoger el resto de (P) de tal forma que (G) y (P) en su conjunto sean compatibles con la experiencia. (Einstein 1922, 35)

Por un lado, merece la pena mencionar que a pesar de que Einstein atribuye este punto de vista a Poincaré, tal posición parece más bien corresponderse con el convencionalismo holista de

[2] Traducción del autor, basada en la traducción inglesa de 1922.

Duhem (Howard 1990). Pero en cualquier caso, los comentarios de Einstein acerca de la inseparabilidad de física y geometría son importantes porque aclaran el por qué de la importancia atribuida por Cabrera al error experimental, y ponen de manifiesto el carácter convencionalista de su argumento. Desde esta óptica, el problema detectado por Cabrera no sería simplemente que la verdadera geometría del mundo sólo puede conocerse de manera aproximada, dentro de los límites fijados por el error experimental, sino que el proyecto de determinar empíricamente tal geometría exige un tratamiento sistemático el error experimental. Pero dado que todo cuerpo físico está sujeto a fenómenos de dilatación y deformación de varios tipos, un tratamiento sistemático del error experimental exige que las leyes físicas que gobiernan dichos procesos sean especificadas. Y como el propio Einstein concluye, se sigue de esto que diferentes combinaciones de geometría y física han de ser compatibles con la experiencia, y que la elección entre las mismas se convierte así en una cuestión de convención.

Conclusión

Como hemos visto, la influencia del convencionalismo de Duhem en el pensamiento de Cabrera es notoria en su análisis de la evidencia empírica a favor de la relatividad y en sus puntos de vista acerca de la relación entre física y geometría. Parece razonable concluir, por lo tanto, que que la sofisticación filosófica de Cabrera es mayor de lo que se ha venido suponiendo, y que entre sus actividades como introductor de la relatividad en España debe incluirse su defensa filosófica de la validez de la teoría de Einstein.

Aunque un análisis más en profundidad de la filosofía de la ciencia de Blas Cabrera está aun por hacer, merecería la pena investigar más a fondo su relación con el debate acerca de los fundamentos de la relatividad general que estaba teniendo lugar en el resto de Europa, y jugó un papel clave en el nacimiento de la filosofía de la ciencia contemporánea. En este sentido, son prometedoras sus estancias durante el periodo de 1912 a 1918 en el laboratorio de Pierre Weiss en Zúrich, donde coincidieron algunos de sus principales protagonistas.

Referencias

Cabrera, B. 1923, *Principio de Relatividad*, Madrid, Publicaciones de la Residencia de Estudiantes.

Duhem, P. 1914, *La Théorie Physique: Son Objet et sa Structure,* Paris, Marcel Riviera & Cie; trans. by P. P. Wiener. 1954, *The Aim and Structure of Physical Theory*, Princeton, NJ, Princeton University Press.

Einstein, A. 1922, 'Geometry and experience', in *Sidelights on Relativity,* New York, E. P. Dutton.

Glick, T. F. 1988, *Einstein in Spain,* Princeton, Princeton University Press.

González de la Posada, F. 1994, *Blas Cabrera ante Einstein y la Relatividad,* Madrid, Ibergráficas

S.A.
Howard, D. 1990, 'Einstein and Duhem', *Synthese* 83, 363-84.
Poincaré, H. 1902, *La Science et l'Hypothèse*. Paris, Flammarion; trans. by W. J. Greenstreet. 1905, *Science and Hypothesis*, London, NY, Scott.
Sánchez Ron, J. M. and Roca-Rosell, A. 1993, 'Spain's first school of physics: Blas Cabrera's Laboratorio de Investigaciones Físicas', *Osiris* 8, 127-55.

THREE-DIMENSIONALIST BOHMIAN MECHANICS: DOES NEWTON DEFEAT ARISTOTLE?[1]

ALBERT SOLÉ
Universitat Autònoma de Barcelona
Albert.Sole@uab.cat

ABSTRACT: Bohmian mechanics can be formulated either in a quasi-Newtonian form, with the postulation of forces proportional to acceleration, or in a first-order form that can be interpreted as incorporating causal concepts akin to those of Aristotelian physics. Here I analyze whether the desideratum of an interpretation of Bohmian mechanics both explanatorily adequate and not committed to the existence of the wave function in configuration space favors one approach to the theory over the other. Contrary to some recent claims in the literature I argue that the quasi-Newtonian approach based on the idea of a quantum potential is not favored.
KEYWORDS: Bohmian mechanics; wave function ontology; configuration space; quantum potential; causation; explanation.

1. Introduction

In the literature, two important interpretational debates have arisen among partisans of Bohmian mechanics. The first concerns the status and the nature of the wave function, as interpreted within the theory. As is well known, the wave function of a closed *N*-particle system is *not* defined in three-dimensional space but rather in a 3*N*-dimensional space usually referred to as *configuration space*. This renders an interpretation of the wave function as a real physical field highly problematic and some authors like Dürr *et al.* (1992, 1997) [hereafter, DGZ] avoid this difficulty defending that Bohmian mechanics is only committed to the existence of the Bohmian particles in ordinary three-dimensional space. The second debate concerns the fundamental law of movement of the Bohmian particles and, arguably, the physical symmetries and explanatory resources available to the Bohmian. Thus, according to partisans of the so-called *guidance approach* to the theory, Bohmian mechanics is essentially a first-order theory[2] and can accommodate causal concepts akin to those characteristic of Aristotelian physics. In contrast,

[1] Presented at the 7th Conference of the Spanish Society of the Spanish Society for Logic, Methodology and Philosophy of Science, Universidade de Santiago de Compostela, Spain, July 18–20, 2012.
[2] In a first-order theory, it is the velocity –the first-order derivative of the position– rather than the acceleration that has a privileged role.

partisans of the so-called *quantum potential approach* defend that Bohmian mechanics is essentially a second-order theory and that concepts such as 'force' and 'potential,' drawn from classical Newtonian mechanics, play a fundamental role.

Although these debates seem to be completely orthogonal, recent developments in the literature apparently reveal an unexpected connection. As mentioned above, DGZ are the first relevant interpreters of Bohmian mechanics that explicitly deny the existence of the wave function as a physical thing, or any other entity defined in configuration space. In what follows, I will refer to interpretations with such a commitment as 3-dimensionalist interpretations of the theory. In addition, DGZ are also explicit in rejecting second-order dynamical concepts and favoring the guidance view. This interpretation has, however, been dismissed as explanatorily wanting by Belousek (2003) and Suárez (2007). Both Belousek and Suárez reply to DGZ by proposing alternative interpretations of Bohmian mechanics that –while keeping with the 3-dimensionalist desideratum– do better than DGZ in explanatory terms. It turns out that both Belousek's and Suárez's proposals draw on resources characteristic of the quantum potential approach. Thus, a natural moral to derive is that this latter approach has better prospects when it comes to accommodating an interpretation of Bohmian mechanics both explanatorily adequate and not committed to the existence of entities in configuration space. In this brief paper, I argue against this conclusion by outlining two new 3-dimensionalist interpretations of the theory that are as explanatory as Belousek's and Suarez's own interpretive proposals and draw only on resources characteristic of the guidance approach.

2. Bohmian mechanics: the guidance vs. the quantum potential approach

The minimal content of Bohmian mechanics can be captured by means of four postulates. The first postulate states that a complete characterization of a closed physical system of *N* particles is given through the specification of the positions of the particles and the wave function of the system. As we have already indicated, the wave function is defined in configuration space, that is, it has the form $\Psi(q,t)$, where. The second postulate of the theory sanctions that the dynamics of the wave function is given by the Schrödinger equation:

(SE) $$i\hbar\frac{\partial\Psi(q,t)}{\partial t}=\hat{\mathrm{H}}\Psi(q,t)$$

where $\hat{\mathrm{H}}$ is the Hamiltonian operator of the system. According to the third postulate, writing the wave function in polar form $\Psi(q,t)=\mathrm{R}(q,t)\mathrm{e}^{i\mathrm{S}(q,t)/\hbar}$, the temporal evolution of the particles'

configuration depends on the wave function of the system and is given by the so-called *guidance equation*:

(GE) $\mathrm{m}_k \vec{\mathrm{v}}_k = \vec{\nabla}_k \mathrm{S}(q,t)$

where $\vec{\mathrm{v}}_k$ is the velocity of the *k*th particle and m_k is its mass. Finally, the statistical postulate ensures that, at given instant *t,* the epistemic probability for the configuration *Q* is given by the usual Born's rule $\rho(Q,t) = |\Psi(Q,t)|^2$.

Now, whereas advocates of the guidance view take the content of Bohmian mechanics to be exhausted by the postulates above and, consequently, assume the guidance relation (GE) to be the fundamental law for the particles' evolution, advocates of the quantum potential approach defend that the fundamental dynamics of the Bohmian particles is given by the *Modified Newton's Second Law:*

(MNSL) $\mathrm{m}_k \vec{\mathrm{a}}_k = -\vec{\nabla}_k (\mathrm{U}(q,t) + \mathrm{V}(q,t))$

Notice that this equation is identical to Newton's Second Law except for the introduction of the so-called «*quantum potential*»,

(QP) $\mathrm{U}(q,t) \equiv -\sum_k \left(\frac{\hbar^2}{2\mathrm{m}_k} \right) \frac{\vec{\nabla}_k{}^2 \mathrm{R}(q,t)}{\mathrm{R}(q,t)}$

that gives rise to a genuinely quantum force.

All solutions of the (GE) are also solutions of the (MNSL). But because it is second-order, this latter equation admits many more solutions than the former. Therefore, in order to ensure consistency with the statistical postulate, advocates of the quantum potential approach have to get rid of the surplus content afforded by (MNSL) and they do so by restricting the initial velocities in such a way that they conform to the guiding relation at time t_0,

(1) $\vec{p}_{k0} = \mathrm{m}_k \vec{\mathrm{v}}_{k0} = \vec{\nabla}_k \mathrm{S}(q(t_0), t_0)$.

In consequence, (GE) still has a role within the quantum potential approach, but it is interpreted *not* as a fundamental dynamical equation but as a selection rule constraining admissible initial conditions.

The whole point of the quantum potential approach is to import the usual explanatory resources of classical Newtonian mechanics. This is clearly so because *via* (MNSL) the twists and bounces of the Bohmian particles can naturally be interpreted as being caused by forces proportional to the acceleration, energy balances can be used in the relevant explanations, etc. It has

to be noted, however, that Bohmian phenomena do not conform exactly to a Newtonian mold, for the quantum forces arising from the quantum potential are extremely non-classical. These forces ultimately depend on the wave function and, consequently, they are multi-particle rather two-particle and they are not a pre-assigned function of the inter-particle distances.

In contrast, it has been noted that the guidance approach can underpin an explanatory and causal framework in a sense reminiscent of Aristotelian physics. Undoubtedly, Valentini is the author that has pushed the Aristotelian interpretation of Bohmian mechanics furthest, with his 1997 defense that the vectors '$\vec{\nabla}_k \mathrm{S}(q,t)$' appearing on the right hand side of (GE) can consistently be understood as Aristotelian 'forces' that are proportional not to the acceleration but to the velocity. Thus, for Valentini, (GE) has a genuinely dynamical content and, having their origin in the wave function, the Aristotelian 'forces' $\vec{\nabla}_k \mathrm{S}(q,t)$ must be regarded as causes of the motion of the Bohmian particles.

This very idea of having Aristotelian 'forces' in a theory that, after all, is Galilean invariant seems blatantly contradictory. Valentini fully appreciates the contradiction, but, far from giving up his interpretation, he argues that the alleged Galilean symmetry of Bohmian mechanics is *not* fundamental but a fictitious symmetry arising from the unnoticed introduction of fictitious Aristotelian 'forces.' Given the usual way the wave function is taken to transform under a Galilean boost, it is easy to show that if Σ' is a coordinate reference system moving away from system Σ at a constant velocity $\vec{\mathrm{V}} \equiv (\mathrm{V}_1, \mathrm{V}_2, \mathrm{V}_3)$, then the following relation obtains:

(2) $\vec{\nabla}'_k \mathrm{S}' = \vec{\nabla}_k \mathrm{S} - \mathrm{m}_k \vec{\mathrm{V}}$.

Valentini considers that the transformation above does not reflect a real symmetry of Bohmian mechanics and interprets the term '$-\mathrm{m}_k \vec{\mathrm{V}}$' as a fictitious Aristotelian 'force' whose introduction preserves the applicability of (GE) in the primed reference system Σ'. In this regard, the situation would be analogous to that in which fictitious forces like the Coriolis force are introduced in classical Newtonian mechanics order to apply Newton's laws in non-inertial terrestrial laboratory frames. In the classical case, however, we can easily identify fictitious forces for we have a clear agreement on the origin and nature of forces and, consequently, a standard for unforced free motion. This would not be the case in Bohmian mechanics, because both the Aristotelian 'forces' appearing on the right hand side of (GE) and the Newtonian forces appearing in the right hand side of (MNSL) are not pre-assigned functions of inter-particular distances and do not generally fall off

with these distances. To sum up, according to Valentini's interpretation of Bohmian mechanics, although the theory sanctions that there is a state of absolute rest, it cannot be effectively detected.

It is worth noticing, finally, that if a view like Valentini's is to be accepted, then the differences between the guidance and the quantum potential approach to Bohmian mechanics would include not only the fundamental equations of motion of the Bohmian particles and the explanatory resources available, but also the space-time structure and the symmetries posited by the theory.

3. DGZ and their critics

DGZ disagree with Valentini in assuming that Galilean invariance is a fundamental symmetry of Bohmian mechanics. As pointed out before, the former authors interpret Bohmian mechanics as postulating the existence of no physical entity over and above the Bohmian particles. These particles are taken to obey the first-order (GE) and dynamical concepts such as force, energy, etc., are altogether rejected.

Belousek (2003) appreciates the DGZ interpretation because it does not commit to configuration space realism, but dismisses it as being explanatorily inadequate. In this regard, the author claims (Belousek 2003, 128) that an adequate interpretation of Bohmian mechanics must somehow interpret the wave function in physical (and not merely statistical) terms and that the physical interpretation of the wave function should adequately underwrite explanation of quantum phenomena.

Although Belousek does not explicitly endorse any specific account of explanation, it is clear that he understands «explanation» in causal physical terms. The *explananda* are the Bohmian trajectories themselves and a fully fledged version of the quantum potential approach provides the paradigm of a satisfactory explanation: the movement of the particles is explained in terms of the (Newtonian) forces acting upon them in accordance with (MNSL) and, in turn, the cause of such forces is located in the wave function, interpreted as a real physical field.

It is no surprise that, having this account of explanation in mind, Belousek is dissatisfied with the austere ontology of DGZ: since these authors allow for no physical entity other than the Bohmian particles and reject any resort to dynamical concepts, they have no resources to provide a satisfactory explanation of the particles' trajectories in physical causal terms.[3] But Belousek can neither accept a fully fledged version of the quantum potential approach for it is committed with the existence of the wave function as a physical field in configuration space. Thus, Belousek's aim is to

[3] Here I am merely reporting Belousek's criticism which I do not share completely. For more details, see Solé (forthcoming).

inflate the sober ontology allowed by DGZ in order to be able to tell a better causal/explanatory story while keeping the ontology three-dimensional. As we will see next, in a sense this is also Suárez' aim.

Belousek's own interpretive proposal is dualistic. It consists in postulating that over and above the Bohmian particles there are the three-dimensional forces $-\vec{\nabla}_k\left(\mathrm{U}+\mathrm{V}\right)$ which act upon the particles in a Newtonian manner, producing accelerations, as prescribed by (MNSL). As in this interpretation the existence of the wave function as a real physical field is explicitly denied, the forces are not taken to have their origin in the wave function but are deemed primitive. Belousek admits that this may seem unsatisfactory but recalls that in Newton's theory the gravitational forces are also left without an explanation and claims that, after all, this could be the best possible balance between explanatory power and the requirement of three-dimensionality.

Suárez (2007) suggests inflating the ontology in a different way, closely connected to his defense of a *selective propensities* interpretation of quantum mechanics (see Suárez 2004). In brief, Suárez attributes to every quantum system a full catalogue of propensities, that is, objective and real dispositional properties of the system entailing particular manifestations with specific probabilities if the right test conditions apply. In his 2007, Suárez attempts to show that similar dispositional notions can be profitably applied in other interpretations of quantum mechanics. In the case of Bohmian mechanics, Suárez explicitly resorts to the quantum potential approach and considers the force field in configuration space defined by (MNSL) to describe modalities that, in turn, are interpreted as a full catalogue of the propensities of the Bohmian particles.

It must be noted that as Bohmian mechanics is a deterministic theory, the propensities of the Bohmian particles would be sure-fire dispositions. Like Belousek, Suárez rejects interpreting the wave function as a real physical field but it nevertheless has some indirect physical significance because of its role in determining the particles' objective propensities. These propensities are credited with an essential explanatory role, for the objective propensity of a particle to exhibit a certain manifestation under the right conditions is taken to explain the occurrence of the manifestation itself.

4. Two alternative proposals

Both Belousek's and Suárez' interpretations are explicitly embedded within the quantum potential approach, whereas the DGZ interpretation is the only three-dimensionalist interpretation of Bohmian mechanics within the guidance approach vindicated in the literature so far. Belousek pretends that his own interpretation is more explanatory than the DGZ interpretation because of the

postulate of the primitive (Newtonian) forces appearing in the right hand side of (MNSL) and the availability of all explanatory resources of classical mechanics. Similarly, Suárez considers that attributing a full catalogue of the propensities to the Bohmian particles renders his interpretation superior in explanatory power than that of DGZ. If we agree with Belousek's and Suárez' diagnosis, we may be tempted to conclude that the quest for an interpretation of Bohmian mechanics both three-dimensionalist and explanatorily adequate favors the quantum potential over the guidance approach. I now proceed to show that this conclusion must be resisted by advancing two new interpretations of Bohmian mechanics that parallel the explanatory power of Belousek's and Suárez's proposals, drawing only on resources available to partisans of the guidance approach. The key is to rely *not* on the framework provided by DGZ, but on the dynamical reinterpretation of the guidance approach to Bohmian mechanics in terms of Aristotelian 'forces' developed by Valentini.

My first suggestion consists in mimicking the very same strategy devised by Belousek but using the resources that Valentini has made available. As a result, one is lead to a dualistic interpretation of Bohmian mechanics that postulates the existence of the Bohmian particles and the set of three-dimensional Aristotelian 'forces' $\vec{\nabla}_k \mathrm{S}(q,t)$ that act upon the particles according to (GE). Here, as in Belousek's own interpretation, the wave function is not interpreted as a real field; but it nevertheless retains indirect physical significance for it determines the Aristotelian 'forces' that, in turn, are interpreted as primitive physical causes of the particles' movement. It must be noted, in addition, that this interpretation does not conflate with Valentini's own interpretation, for Valentini commits to configuration space realism.

Suárez (2007) explicitly grounds his interpretation of Bohmian mechanics in terms of propensities in the quantum potential approach to the theory. I think one of his reasons for doing so is that Suárez conflates the guidance view with the austere interpretation of DGZ. But nothing in the very idea of a disposition as attributed to Bohmian particles seem to depend on whether the dynamics of the latter is first-order or second-order. In this respect, if the modalities described by the (Newtonian) force field arising from the quantum potential can be interpreted as a catalogue of the propensities of the Bohmian particles, so the modalities described by the (Aristotelian) 'force' field $\vec{\nabla}\mathrm{S}$ can. This latter interpretation would lead to a propensity interpretation of Bohmian mechanics fully embedded within the guidance approach.

5. Conclusions

So far I have shown that the search for a three-dimensionalist and explanatorily adequate interpretation of Bohmian mechanics does not provide reasons for favoring the quantum potential approach to the theory by advancing two interpretations of Bohmian mechanics that parallel the explanatory power of those of Belousek and Suárez and that fully operate within the guidance approach. Given Valentini's (1997) defense of the tenability of an Aristotelian causal and explanatory framework underpinning the guidance approach to Bohmian mechanics, preference for Newton over Aristotle seems to be just a matter of habit or philosophical taste.

I have not committed myself either to Belousek's criticism of the DGZ interpretation nor to Belousek's and Suárez' alternative interpretive proposals being better than DGZ. Neither have I explored whether the two new interpretations outlined here present advantages over Belousek's and Suárez's proposals. This is surely a task worth undertaking, but it is not needed to reach the conclusion that Newton does *not* defeat Aristotle when it comes to accommodating a three-dimensionalist, explanatorily adequate interpretation of Bohmian mechanics.[4]

REFERENCES

Belousek, D.W. 2003, 'Formalism, Ontology and Methodology in Bohmian Mechanics', *Foundations of Science* 8(2), 109–172.
Dürr, D., Goldstein, S., and Zanghì, N. 1992, 'Quantum Equilibrium and the Origin of Absolute Uncertainty', *Journal of Statistical Physics* **67**, 843–907.
_____ 1997, 'Bohmian Mechanics and the Meaning of the Wave Function' in Cohen *et al.* (eds.) *Experimental Metaphysics — Quantum Mechanical Studies for Abner Shimony, Volume One, Boston Studies in the Philosophy of Science* **193**, Boston: Kluwer Academic Publishers, 25−38.
Suárez, M. 2004, 'Quantum Selections, Propensities, and the Problem of Measurement', *British Journal for the Philosophy of Science* 55, 219–255.
_____ 2007, 'Quantum Propensities', *Studies in the History and Philosophy of Modern Physics* 38, 418–438.
Solé, A. (forthcoming), *Bohmian mechanics without wave function ontology*.
Valentini, A. 1997, 'On Galilean and Lorentz Invariance in Pilot-Wave dynamics', *Physics Letters* A 228, 215–222.

[4] See Solé (forthcoming) for a completion of this project.

EL PLANTEAMIENTO SOBRE LA MUJER: POSITIVISTAS, KRAUSISTAS Y CATÓLICOS[1]

Mª JOSÉ TACORONTE DOMÍNGUEZ
Universidad de La Laguna

ABSTRACT: With this participation is to talk with three streams fundametales, positivism, catholicism and krausismo, about the conception of women in Spain in the late nineteenth and early twentieth century.
KEYWORDS: Positivism, Catholicism, Krausismo and Women

El ideal sobre el sexo femenino dio lugar a arduas disputas entre diversas corrientes ideológicas que interpretaban y reinterpretaban la concepción del denominado *sexo débil* en la sociedad finisecular española. La concepción dominante derivaba de un pensamiento profundamente naturalista y cientificista, que llegó al contexto español de la mano del positivismo europeo. Nos centraremos en mostrar las disquisiciones sobre la cuestión de la mujer, qué funciones sociales podía desempeñar, cuál era su finalidad, etcétera, desde el acercamiento a tres posturas fundamentales: positivistas, krausistas y católicos.

El debate intelectual que se instauró en España entre los siglos XIX y XX se caracterizó por la influencia de corrientes científicas y filosóficas importantes en la Europa del momento. Los avances de la modernidad unidos a los cambios económicos, al surgimiento de la clase social burguesa y al empuje de la industrialización, de forma externa, junto al fracaso del sexenio revolucionario, la crisis económica y la restauración borbónica de forma interna, hicieron de este periodo un campo de cultivo muy interesante para el estudio de los ideales de género en el territorio español.

La tensión es clara entre las tendencias modernizadoras y las tradicionalistas, entre las primeras destacan el positivismo y el krausismo, y entre las segundas el catolicismo. Estas corrientes constituyeron un marco de gran calado para el estudio de la concepción de las mujeres y el pensamiento feminista.

[1] La redacción de este trabajo ha sido facilitada por la participación en el proyecto de investigación, *Epistemología, política e institucionalización en el desarrollo científico: La ciencia española de la república a la dictadura.* FFI2009-09483/FISO. Y por la cobertura de la Beca para Postgraduados de CajaCanarias-Universidad de La Laguna.

Las ciencias biológicas y el discurso médico en especial, fueron centrales en la configuración de los ideales de género en este contexto. A ellos también se sumó la influencia del pensamiento krausista y de la ortodoxia católica, pensamiento, éste último, enraizado en todo el territorio español.

El debate acerca de la cuestión femenina se enmarcó en la polémica entre ciencia y religión, o lo que es lo mismo, entre *renovación y continuidad*. La ciencia se convirtió en la medida de todas las cosas, dando paso a planteamientos naturalistas y evolucionistas. Se aceptó la antropología física, el evolucionismo social y la psicología diferencial que provenían de otros países europeos, introduciéndose posteriormente en España.[2] Grosso modo, estas teorías enfatizaban las diferencias entre los sexos y las razas humanas, aplicando las leyes naturales al análisis social. El destino, pues, lo marca la biología y estas novedosas tendencias justificaban mediante la fisiología, la frenología y la craneología la inferioridad mental y física de la mujer. Por la contra, los intelectuales católicos reaccionaron con cierta virulencia contra estos planteamientos dando lugar a un debate a veces encendido. Los positivistas, aunque apuestan por la idea de progreso con apoyo en los avances científicos, sin embargo, respecto a la mujer van a mantener una posición conservadora, custodiando el orden sexual existente. La mujer es presentada como rémora del progreso, pasando a constituir un objeto de estudio científico.

Por otro lado, los pensadores católicos, más conservadores y/o tradicionalistas tenían una predisposición más favorable a la educación de las mujeres (sacarlas de su situación de ignorancia), aunque mantenían las diferencias y las dicotomías entre los roles sociales de los sexos. Consideraban que las mujeres no estaban determinadas biológicamente, y en su cuerpo de doctrinas entendían que la inteligencia y el espíritu no tienen sexo. Todo ello le valió a la iglesia para apoyar, en cierta manera la causa femenina y criticar los posicionamientos positivistas. Estas ideas de talante progresista se pueden entender como una estrategia de la iglesia para mantener su control ante los fieles, que en su mayoría eran mujeres, como se explicará seguidamente.

Y en tercer lugar, los krausistas constituían un punto intermedio entre estos dos polos contrapuestos. Desde su filosofía racionalista y humanista, guiados también por su política burguesa reformista, atendían a un cristianismo racional con fe en el progreso, mediante la ciencia; a la vez que mantenían la idea de la educación como capacidad transformadora de la sociedad. Defendían la educación femenina y la mejora de sus condiciones sociales. Y a diferencia de los positivistas, no recurrieron a leyes naturales inamovibles, ni tampoco siguieron la estela elitista del cristianismo, eran igualitaristas. Para estos pensadores, continuando la huella de J.S. Mill y el

[2] Para un desarrollo más pormenorizado véase Gómez, 2003.

ambientalismo, no hay diferencias innatas entre los seres humanos, y sí una ilimitada posibilidad de perfeccionamiento por medio de la educación.

La bipolaridad y tensión la mantienen, sobre todo, católicos y positivistas. Para los primeros las pretensiones de los científicos eran desmesuradas y arrogantes. Mientras que para los segundos, la ciencia es la clave que permite explicar toda la naturaleza, incluyendo la humana, para desvelar así sus misterios. Los fervientes hombres de ciencia, de talante progresista y democrático, chocan con cualquier intento de igualdad entre hombres y mujeres; a diferencia de la mejora educativa que defienden los católicos y la igualdad social y educativa, de talante más moderno que reivindican los krausistas.

1. La mujer desde la concepción católica: avance para conservar

El discurso católico de género sigue unos patrones continuos que podemos encontrar en obras que trascendieron su propia época, como ejemplo, *El ángel del hogar* de Pilar Sinués de Marco, las obras de Severo Catalina, o las del padre Antonio Mª Claret. En ellas observamos que el arquetipo femenino que defiende y difunde el pensamiento católico entiende a la mujer como ser abnegado, paciente, cándido, generoso, tierno, y a la vez, es un ser de gran fragilidad y sensibilidad. La feminidad se define por *ser otro, para el otro, a través del otro*. Las funciones que desempeña la mujer son las del hogar y la maternidad. Marcada pues por una instrucción deficiente, la mujer es presentada como ferviente católica, sin preocupaciones culturales ni políticas, y al margen de la vida profesional. Se sobreentendía que la carrera de la mujer es el matrimonio y su labor como ángel del hogar se orienta en exclusiva a la crianza y al cuidado de la familia. La ortodoxia sobre la mujer puede resumirse en una de las frases de Severo Catalina: “La mujer que no está organizada para amar, no es mujer” (Catalina 1951, 49).

La relación jerárquica que se establece entre los sexos configura un *status quo* sexual totalmente favorable para los varones, fundamentado en la división clásica de los roles de género. La mujer es arrojada a situaciones de dependencia, fragilidad y pobreza, cuya única salvación era el matrimonio y la religión.

Esta concepción pasiva sobre la mujer comienza a sufrir ligeros cambios con la influencia del proceso de secularización que se intenta dar en España a finales del S. XIX proveniente de Europa. Este transcurso moderno supuso un abandono masculino de la fe para pasar a engrosar las filas de otro tipo de movimientos y organizaciones sociales, lo cual deja como resultado una feligresía católica totalmente feminizada. Estos cambios suponían una visión más igualitarista de las mujeres, reivindicando mejoras sociales y legales para éstas. Obviamente todo ello no era del gusto de la iglesia, dado que cuestionaba su ideario acerca de lo femenino. Esta situación obligó a la

iglesia a modificar su discurso de género; se reconfigura la concepción de la mujer para evitar la pérdida de fieles, se postula *la pastoral de la mujer*; un nuevo ideario que representó una reconversión atenuada de la imagen de la mujer, para así poder contrarrestar las fuerzas secularizadoras que podían operar sobre ella. Estos cambios presentan al catolicismo con una religiosidad más sentimental y delicada, más benevolente y protectora, resaltando la idea de que las mujeres se caracterizan por tener una moral superior. Todo ello facilita un discurso más mitigado en comparación con el discurso de la domesticidad clásico, a la vez que abre la posibilidad a un nuevo campo de actuación para las mujeres, entendidas no sólo como esposas y madres, sino ahora también como mujeres sociales. Con estas modificaciones se hacían extensivas las actividades propias de su sexo, ahora en el ámbito público, social. Se pasa pues, de la idea de Eva pecadora arraigada en el Antiguo Régimen, a la visión de María redentora. Las mujeres son habilitadas para participar de forma concreta en la vida pública, serán las encargadas de recristianizar a la sociedad. En este sentido el ideario católico no desdeña la instrucción de las mujeres, sobre todo por dos razones estratégicas fundamentales:

1.-porque la mujer era la encargada de educar a la progenie, pudiendo extender la palabra de dios y 2.- porque esta educación mantenía el control y la influencia que la iglesia sustentaba en un país profundamente católico, como era el caso de España.

Esta ligera adaptación de la iglesia es más mitigada en comparación al discurso de la domesticidad tradicional, porque abre la posibilidad para un nuevo campo de actuación para las mujeres, sobre todo burguesas y aristócratas, en tanto que mujeres sociales, las cuales podían hacer extensivas las actividades propias de su sexo, ahora, en el ámbito público, aunque se ha de resaltar que su trabajo era de carácter caritativo, fundamentalmente conformación de comedores de beneficencia, escuelas para la educación de las clases obreras, asilos, hospitales e instrucción católica. Es evidente que se perpetúa el ideal clásico de género, dividiendo entre las dos esferas y manteniendo la ética del cuidad, pero la incorporación de la pastoral de la mujer y el hecho de contemplar a la mujer como portadora de una moral superior hace que se pase del simbolismo de la *Eva pecadora*, donde la mujer es la culpable de los males sociales, para entenderla ahora como descendiente de la *María redentora*.

2. El krausismo y la educación femenina.

En el contexto intelectual e ideológico decimonónico y de principios del XX, el pensamiento krausista tuvo un protagonismo considerable en varios terrenos, pero se ha de destacar, sobre todo, el educativo. Fueron los intelectuales krausistas los que abogaron por una igualdad educativa entre los sexos, constituyendo la perspectiva más igualitarista de las tres que destacamos. La situación de

la mujer es muy desfavorable, hay un alto índice de analfabetismo social, donde el grueso corresponde al denominado *sexo débil*. Los pensadores krausistas apostaban por una renovación fuerte de la educación social, y en general de la femenina. Sus miras estaban puestas en el cambio, no sólo en la metodología de la enseñanza, como educadores humanistas, sino también en la práctica educativa y en el contenido de la instrucción que debía transmitirse, y en la cual debían estar formadas las nuevas generaciones.

Mantienen un enfoque radical al defender la educación de las niñas de manera básica e integral. Con el horizonte de que ambos sexos deberían tener una educación igualitaria para poder ejercer las distintas profesiones sin ningún tipo de cortapisas. Así pues, el ímpetu krausista dejó huella en la trayectoria pedagógica y educativa de las mujeres. Fueron los primeros que claramente definieron, según las palabras de Fernando de Castro que:

> "la mujer tiene igual derecho e idéntico deber que el hombre a instruirse para que le sea posible realizar la misión que como individuo se le ha asignado y lo haga con las mayores garantías de éxito". (Capel 1986,120).

Como inicio del apoyo krausista a la causa de las mujeres se ha de destacar *Las conferencias dominicales*, dirigidas por el propio F. de Castro en la complutense de Madrid en 1869. La estrategia de las conferencias estaba orientada a, en palabras de Geraldine Scanlon:

> "convencer a la gente de que la mujer tiene que recibir una educación más extensa si se quiere que cumpla su misión en la vida (…)" (Scanlon 1986, 31).

Este primer conato a favor de la educación de la mujer fue seguido por la Asociación para la enseñanza de la mujer (1870). Ambas empresas destacarán gracias a su ampliación de contenidos en enseñanzas científicas y en mejora pedagógica, lo que implica un alto compromiso con la causa educativa de las mujeres, sin precedentes hasta el momento.

Su orientación no era preparar a la mujer para que fuera mejor madre, sino prepararla para que fuera autosuficiente, y pudiera valerse por sí misma. El discurso krausista fue calando de forma paulatina, pero progresiva. La importancia de la educación entre este grupo de pensadores dio lugar a la creación de diferentes escuelas orientadas a la formación competente de mujeres para su desarrollo profesional desde la enseñanza científica y las artes liberales. Se crean la escuela de institutrices (1871), la escuela de comercio (1878) y la de correos y telégrafos (1883). Detalle a subrayar es que la asociación para la enseñanza de la mujer, así como las diferentes escuelas para mujeres, se crean antes que la mítica Residencia de estudiantes, fundada en 1910. Así pues, los krausistas comienzan a denunciar la ínfima y pésima educación femenina, la necesidad de cambio de su situación, comienzan a reivindicar una educación obligatoria para ambos sexos, la

coeducación no limitada a los niveles elementales, sino también en la educación superior, así como también abogaban por la incorporación de la mujer al trabajo remunerado y al desempeño de cualquier profesión. Los krausistas entendían que la mujer no es un instrumento para perfeccionar a los hijos y llevar el hogar en tanto que ángeles del mismo, sino como individuos autónomos que tienen que estar instruidos para ser independientes, lo que constituye un bien social en sí mismo.
Estas ideas son índice de cómo ciertas mentalidades iban avanzando hacia el progreso social. Los krausistas constituyen pues, una pequeña minoría frente al grueso social, marcado éste por un férreo conservadurismo, anclado en la división de esferas y en la naturalización de los roles sexuales. Las aspiraciones krausistas constituyen ideas pioneras que siembran dudas en una selecta minoría, lo que dista mucho de producir cambios sociales y estructurales significativos, cuestiones que por otro lado, sí consiguió la ciencia.

3. El ideal científico de mujer: conservación para el progreso.

Los avances científicos y su implosión en el contexto europeo tenían el reconocimiento de todo círculo intelectual, la ciencia se convierte en la medida de todas las cosas; y en el caso de las mujeres, el desarrollo científico lo que hizo fue afirma y justificar positivamente lo que la tradición había implantado, la mujer es un hombre imperfecto, y por tanto, inferior.
La concepción de la mujer desde el ámbito científico tiene reminiscencias tradicionales y conservadoras, reformuladas desde la óptica científica, la cual, pretendía explicarlo todo. Grosso modo, la mujer se presenta como diferente al hombre, y la diferencia se entiende como inferioridad. Se rechaza la idea de la pastoral de la mujer donde éstas son portadoras de una moral superior, para pasar a entenderlas como animales incapaces de llegar a la esfera del conocimiento científico. Médicos, positivistas, evolucionistas y científicos tendrán un papel fundamental en el tratamiento de la naturaleza femenina. La influencia provino principalmente de Europa, siendo uno de los pioneros en este terreno Moebius, aunque también se ha de destacar a Max Nordau, Lombroso u Otto Weininger. Mientras que los condiscípulos españoles son Edmundo González Blanco y Roberto Novoa Santos. Todos estos científicos compartían la idea de que las diferencias anatómicas y cerebrales entre hombres y mujeres eran claves para explicar la inferioridad femenina. Así pues, los trabajos de estos hombres de ciencia van encaminados a demostrar, mediante los hechos y los datos, la indigencia espiritual y mental de las mujeres.

El peso del cerebro y su forma constituyeron un fuerte argumento en detrimento de la inteligencia femenina; el tamaño del cerebro indicaba la capacidad mental, lo cual siguiendo un paralelismo entre fuerza y masa muscular, se deducía que un cerebro mayor era indicativo de mayor inteligencia… La relación entre la función sexual y la actividad cerebral, donde el desarrollo de la

primera imposibilita para la actuación de la segunda. La menstruación también fue blanco de los comentarios cientificistas, el flujo menstrual incapacitaba a la mujer para actividades de mayor rigor ya que su estado anímico era más voluble y ello, supuestamente no le permitía atender con claridad a cuestiones intelectuales o de gran rigor. Asimismo, el sistema nervioso fue otro argumento propuesto en detrimento de la inteligencia femenina, siendo éste en la mujer más sensible y por tanto más vulnerable al entorno, lo que justificaba su permanencia (impuesta) en el ámbito privado. En el desarrollo del siglo XX también se originaron cotas de inferioridad, los desarrollos de la bioquímica y la endocrinología marcaban las diferencias entre ambos sexos. El estudio del tiroides y de los principios arsenicales justificó la idea de que el organismo tenía una cantidad limita de energía que en la mujer se emplea en la menstruación y la reproducción, y en el hombre influían en la actividad cerebral.

Para dar por concluido este esbozo, cabe destacar, como síntesis, que lo comentado constituye un acercamiento a un contexto atractivo en el cual fluctúan ciertas paradojas interesantes de poner sobre el tapete:

1.- Los que apuestan por el progreso y el avance científico, es decir, los partidarios del positivismo, buscan justificaciones para apear a la mujer del avance humano, incapacitándola para el desarrollo intelectual, científico y, entre otros, educativo.

2.- Por otro lado, observamos cómo las ideas más conservadoras de la iglesia, en contraposición al avance de la modernidad, apuestan por la educación de las mujeres para mantener su control social, no perder más feligresía, y servirse de ésta además, en tanto que mujeres creyentes, para re-cristianizar la sociedad y a las nuevas generaciones.

3.-Por su parte, los krausistas sientan un precedente de apoyo y de apuesta por el cambio y la mejora de la situación de las mujeres, siendo más acordes con los ideales de modernidad, cambio y liberalismo que parecían *asomarse tímidamente* a España desde Europa.

Por todo ello, de las tres corrientes comentadas, la positivista fue la que imperó debido principalmente a la importancia otorgada a la ciencia y al progreso. La idea de mujer quedó reducida a las ideas de inferioridad y dependencia. La cuestión de la división de las dos esferas (público/privada) defendida también por la iglesia se radicalizó con la visión y las justificaciones positivistas sobre la mujer, ésta es diferente al hombre, y la diferencia la hace inferior, tanto mental como físicamente.

REFERENCIAS

Arce Pinedo, R. 2007, *Dios, patria y hogar. La construcción de la mujer española por el catolicismo y las derechas en el primer tercio del siglo XX,* Santander, Ediciones de la Universidad de Cantabria.

Aresti, N. 1998, "Pensamiento científico y género en el primer tercio del siglo XX", *Vasconia*, 25, 53-72.

Catalina, S. 1951, *La mujer. Apuntes para un libro*, Buenos Aires, Colección Austral. 1ªEd. 1858.

Capel, R.Mª. 1986, "La apertura del horizonte cultural femenino: Fernando de Castro y los congresos pedagógicos del siglo XIX. En VV. AA, 1986, *Mujer y sociedad en España (1700-1975)* Ministerio de Cultura, Estudios sobre la mujer. Madrid. 1ª edición 1982.

Díaz, E. 1973, *La filosofía social del krausismo español. Editorial cuadernos para el diálogo*. Madrid, Edicusa.

Gómez Rodríguez, A. 2003, *Filosofía y metodología de las ciencias sociales*, Madrid, Alianza Editorial.

Nordau, M. 1901, *Psico-fisiología del genio y del talento*. Madrid, Sáenz de Jubera.

Novoa Santos, R. 1908, *La indigencia espiritual del sexo femenino. Las pruebas anatómicas, fisiológicas y psicológicas de la pobreza mental de la mujer. Su explicación biológica.* Madrid. Sempere y compañía Ed.

Scanlon, G. 1986, *La polémica feminista en la España contemporánea (1868-1974),* Madrid, Ed. Akal.

SCIENCE, TECNOLOGY AND SOCIETY

CIENCIA, TECNOLOGÍA Y SOCIEDAD

LA NUEVA LEY DE LA CIENCIA EN ESPAÑA: UNA APROXIMACIÓN CRÍTICA AL CONTRATO SOCIAL SUBYACENTE[1]

IRENE DÍAZ GARCÍA
Ciemat (Spain)
irene.diaz@ciemat.es

ABSTRACT: Science policy defines and frames the relation between science, technology and society, which has been a point of interest in STS studies since its conformation as an academic field. This work critically explores the traditional conception about science policy and the social contract for science, to review its most important critiques and analyze the evolution of Spanish scientific policy. Finally, presents the most relevant consequences of the last law in this sense: the new *ley de la ciencia*, 2011.
KEYWORDS: Scientific policy, social contract, science and technology system, Science, Tecnology and Society.

1. Política científica y contrato social

Determinar qué es exactamente la política científica y cuáles sus modalidades excede ampliamente este trabajo. Con todo, es un lugar común en la literatura especializada situar sus bases en el informe del ingeniero estadounidense Vannevar Bush (1945) *Science, the endless frontier*, donde se encontrarán las claves para la comprensión de los derroteros políticos y el modelo subyacente del denominado "primer contrato social para la ciencia".

Medio siglo más tarde, las críticas a este modelo han ido acumulándose. Entre las más relevantes se encuentran los "mitos" definidos por Sarewitz (1996)[2], así como la aspiración de un nuevo papel para la sociedad, tanto en la determinación de intereses y líneas de investigación, como en la gestión de productos y resultados del proceso. Así, se ha venido solicitando el establecimiento de un nuevo y segundo contrato social[3] que dé cuenta de las insuficiencias del viejo modelo, corrija sus problemáticas y autorice un nuevo rol para la ciudadanía en la gobernanza del sistema

1 Este trabajo ha sido desarrollado con el apoyo del Proyecto "Políticas de la cultura científica" (FFI2011-24582).

2 "Del beneficio infinito" (modelo lineal); "de la investigación sin trabas" (cualquier incremento del conocimiento puede producir beneficio social); "de la rendición de cuentas" (los controles internos garantizan la calidad de la investigación); "de la autoridad" (la ciencia es base objetiva para decisiones políticas) y "de la frontera sin fin" (ciencia y tecnología son valorativamente neutrales). Igualmente, se podrían haber tomado los siete tópicos descritos por Sanmartín (1990), si bien, al poner mayor atención al equivalente tecnológico, se consideran menos adecuados.

3 Algunos autores sostendrían que sería, en realidad, el tercero (Elzinga y Jamison 1995).

tecnocientífico. Aunque la Declaración de Budapest (UNESCO 1999) podría considerarse un intento de establecer ese nuevo contrato social (Sanz Merino 2005), su influencia real ha sido bastante limitada.

El presente trabajo es un examen crítico de la política científica española reciente a través su cristalización en algún tipo de documento, especialmente, la nueva ley de la ciencia (Gobierno de España 2011). El objetivo es identificar la presencia de los viejos mitos en los nuevos documentos, así como determinar hasta qué punto estos textos recogen las reivindicaciones de los estudios CTS. Para profundizar en el análisis, resulta oportuno dar cuenta, además, de algunos rasgos que, si bien están presentes en el informe Bush, no quedan suficientemente destacados en los cinco mitos. En este sentido subyace que:

-existe una concepción triunfalista de la ciencia, de la que se sigue que los Estados han de concederle una atención prioritaria y una financiación igualmente preferente,

-ha de garantizarse el reclutamiento de los mejores cerebros,

-la ciencia es fundamental para la supremacía militar y el prestigio nacional,

-ha de promoverse el comunitarismo e intercambio científicos,

-aunque no sean prioritarios, otros campos académicos han de ser promovidos,

-no queda espacio para la ciudadanía, más allá de nutrir con los mejores cerebros los centros de investigación y recibir los beneficios del desarrollo científico y tecnológico.

2. De dónde venimos: política científica española del siglo XX

En nuestro país, la política científica ha estado plagada de avatares políticos y económicos que han condicionado profundamente su desarrollo. Mientras a mediados del siglo XX EE.UU. y otros países de nuestro entorno adoptaban una política desarrollista y organizaban un verdadero sistema de ciencia y tecnología, nuestro país abandonaba su "edad de plata" a consecuencia de la Guerra Civil y del posterior exilio (interior o exterior) de sus intelectuales más destacados. El resultado será que, a nivel público y privado, la ciencia y tecnología españolas quedarán lastradas por un desarrollo tenue, modelos de actuación ineficaces e, incluso, un tímido reconocimiento de su importancia política y económica (Muñoz 1998). No será hasta los 60 cuando España, incapaz de mantener el estado de autarquía que la había llevado a un completo aislamiento internacional, comenzase a abrirse al exterior y estimulase un primer intento por organizar un sistema que permitiese equilibrar una balanza de dependencia tecnológica claramente desfavorable.

La Transición llegará con una inversión y recursos humanos para investigación en los puestos de cola de la OCDE, pero al menos la ciencia entra en la agenda política, fenómeno de capital importancia para la reestructuración del sistema español de ciencia y tecnología en los años

siguientes (López y Santesmases 2006). Con todo, la herencia recibida condicionará la primera formalización, ya en los 80, de una política pública explícita de ciencia y tecnología en España: la conocida como "ley de la ciencia" de 1986.

¿A qué debía hacer frente dicha ley? Tras dos informes del Senado, y de acuerdo con Emilio Muñoz (1998) -entonces Director General de Política Científica- cabía destacar los siguientes aspectos: 1) escasa relevancia de la I+D para el desarrollo tecnológico y económico del país; 2) insuficiencia del sistema en recursos humanos e inversiones; 3) escasa coordinación del sistema y ausencia de una orientación general; 4) escasez de recursos de los OPIs; 5) contribución prácticamente inexistente del sector privado; y 6) reparto desigual del sistema en términos territoriales.

Con el primer gobierno socialista del siglo XXI, la ciencia toma un nuevo impulso, el presupuesto se incrementa significativamente y se establecen nuevas herramientas políticas para su fomento y regulación (Muñoz y Sebastián 2008). Dado que en los 90 la mayor parte de los países de nuestro entorno habían reestructurado sus regulaciones para adaptarse a la nueva estructura política de Europa, la globalización, el desarrollo de la sociedad del conocimiento y las nuevas demandas sociales (García Arroyo *et al.* 2007), se generaliza la idea de que es necesaria una nueva ley para la ciencia y tecnología españolas y se comienza su desarrollo. Para cuando llega al Congreso, la actual crisis económica había dado lugar ya a un nuevo clima de recortes.

3. A dónde vamos: política científica española para el siglo XXI

Para tratar de responder a la cuestión de hasta qué punto esta ley reciente responde a los cambios operados a nivel social, académico y político y a si persisten indicios que delaten la presencia del viejo modelo, tomando como referencia el informe Bush (1945) se han examinado la ley de 2011 y la primera comparecencia en el Congreso de la entonces flamante Ministra de Ciencia e Innovación, Cristina Garmendia. Si la nueva ley de la ciencia plasma las aspiraciones políticas de la Cámara para la ciencia y la tecnología españolas en un futuro próximo, la comparecencia recoge las expectativas del Ministerio, quizá más ambiciosas, previas a su trámite parlamentario. Además, se han revisado un texto del entonces Secretario de Estado de Universidades e Investigación, Miguel Ángel Quintanilla (2007) y una entrevista que concedió poco después de abandonar el cargo.

Tanto Quintanilla (2007) como la Ministra y la propia ley parecen tener una concepción economicista del conocimiento, considerado como "*una de las fuentes principales de la riqueza y del bienestar* social" (Quintanilla 2007, 184). Este aspecto requiere especial atención porque enlaza con el primero de los mitos de Sarewitz (1996) y el modelo lineal de innovación que, pese a haber

sido ampliamente criticado por su carácter simplista, parece continuar siendo la concepción subyacente: hemos de promover la ciencia porque será el motor que impulse el crecimiento económico y, con él, la mejora social. Así, Garmendia afirmará que

> "contamos con un amplio consenso social y político en torno a la necesidad de **hacer del conocimiento el eje central del desarrollo social y económico del país** […] desde que este se genera […] hasta que reporta un beneficio a la sociedad" (MICINN 2008),

idea que recoge igualmente el Preámbulo de la ley:

> "la generación de conocimiento en todos los ámbitos, su difusión y su aplicación para la obtención de beneficio social o económico" son fundamentales para el país (Gobierno de España 2011).

En cuanto al mito de la investigación sin trabas, sigue aceptándose la autonomía del científico para determinar sus focos de interés, aunque se les imponen algunas modificaciones importantes como atenerse a "las prácticas y los principios éticos reconocidos", el respeto a la propiedad intelectual o la atención a "las limitaciones presupuestarias o de las infraestructuras"; además de imponérseles, entre otras, la obligación de "procurar que su labor sea relevante para la sociedad" (Artículo 13).

También pueden observarse cambios en lo referido a la rendición de cuentas que, aunque se mantiene dentro de la academia, es objeto de restricciones como la necesidad de incorporar transversalmente el enfoque de género[4], comprometerse con su difusión adoptando mecanismos de acceso abierto[5] o apoyar la cooperación científica (Preámbulo). Además, se indica que la evaluación deberá atender a

> "aspectos científicos, técnicos, sociales, de aplicabilidad industrial, de oportunidad de mercado y de capacidad de transferencia del conocimiento, o cualquier otro considerado estratégico" (Artículo 5).

En paralelo, una nueva orientación de la neutralidad valorativa se podría ver reflejada en la creación del Comité Español de Ética de la Investigación, que queda sin embargo bastante limitado

4 Aunque por espacio no es posible profundizar en la atención al género que específicamente se recoge en esta ley, supone un paso importante a favor del aprovechamiento de todo el capital intelectual, como ya sugería Bush en su informe. En España, los estudios realizados muestran un serio desequilibrio en el número de hombres y mujeres del sistema nacional de ciencia y tecnología, brecha que se incrementa al aumentar el escalafón académico y profe sional y que la ley de 1986 no habría logrado corregir (véase, por ejemplo, Pérez Sedeño y Alcalá 2006).

5 La aspiración de cooperación científica habría sido recuperada, en el marco de la actual sociedad del conocimiento, con el formato de repositorios de acceso abierto (Gobierno de España 2011, Artículo 37); una necesidad considerada imperiosa para una nueva reorganización del sistema español de ciencia y tecnología (Lafuente 2006).

al no especificarse en el texto cuál será su composición (más allá del carácter de *"expertos reconocidos en el ámbito internacional*" de sus integrantes –Artículo 10-), por lo que no es posible precisar si se presupone la existencia de componentes éticos y valorativos en el seno de la institución científico-tecnológica, o bien su actividad se limitará a la redacción de códigos deontológicos.

Finalmente, se mantiene el papel asesor de los expertos, como refleja la creación del Consejo Asesor de Ciencia, Tecnología e Innovación para informar a las Administraciones públicas cuando estas lo requieran (Artículo 9). Más ambicioso en este sentido, aunque igualmente poco concreto, resulta Quintanilla en su entrevista de 2008.

Es posible afirmar, por tanto, que parece darse una persistencia de los cinco mitos definidos por Sarewitz (1996), si bien su presencia queda atenuada por los matices que se han ido señalando. Con respecto a los otros rasgos extraídos a partir del informe de Bush (1945):

-se mantiene una concepción triunfalista de la ciencia, aunque desde una perspectiva menos inocente que encierra el reconocimiento tácito, en su apelación a la sostenibilidad, de que el desarrollo científico-tecnológico ha originado tanto ventajas como algunos problemas.

-el estímulo de las vocaciones científicas y el reclutamiento de los mejores cerebros siguen siendo necesidades urgentes. Quintanilla (2007) focalizará su atención en el desarrollo de mecanismos eficaces para la financiación de los estudios superiores, mientras la nueva ley se centrará en incrementar el atractivo de la carrera profesional y la contratación de jóvenes investigadores (algo que ya contemplaba la Ministra en su comparecencia inicial).

-la ciencia no es ya concebida como empresa nacional (aunque el prestigio parece seguir siendo importante, de acuerdo con el Preámbulo de la ley), sino colaborativa, internacional y globalizada.

-la importancia de otros campos de conocimiento parece haberse incrementado, como revela que la ley sea única para todos ellos, aunque no se ha roto la brecha entre las dos culturas.

-el papel de la ciudadanía sigue siendo escaso. Aunque el Artículo 2 recoge entre sus objetivos

> "promover la participación activa de los ciudadanos en materia de investigación, desarrollo e innovación, y el reconocimiento social de la ciencia a través de la formación científica de la sociedad y de la divulgación científica y tecnológica, así como el reconocimiento de la actividad innovadora y empresarial"

y es un avance respecto a la anteriormente descuidada tarea de divulgación de los investigadores (Artículo 37); el énfasis se pone en potenciar la comunicación y divulgación científicas para el

incremento del apoyo social a la ciencia y no en articular mecanismos concretos para una participación más activa.

Finalmente, se constata la emergencia de la innovación como valor en alza. Se afirmará así, en el Preámbulo de la ley, que "*la apuesta por la innovación es estrictamente necesaria para el crecimiento y competitividad de nuestro sistema productivo*". De modo similar se expresará también el antiguo Secretario de Estado (2008).

Conclusiones: lo viejo y lo nuevo

Como se ha señalado, persisten gran cantidad de paralelismos entre las viejas y nuevas propuestas en política científica. Así, ha sido posible rastrear vestigios de todos los antiguos mitos, si bien es justo señalar que aparecen formulados con importantes matices. Por tanto, o bien la crítica realizada desde los estudios sociales de la ciencia no ha permeado suficientemente en la esfera política, o bien el modelo no está, al fin y al cabo, tan desacertado. En cualquiera de los casos, un ejercicio de autocrítica del propio campo de trabajo (los estudios CTS) parece una tarea necesaria.

Por otra parte, y a pesar de la evolución social, económica y científica de los últimos cincuenta años, los desafíos a los que Bush pretendía responder parecen estar aún en vigor. Así, la política científica sigue teniendo que resolver, al menos, cuestiones como qué hacer para fomentar el desarrollo médico, cómo apoyar la investigación pública y privada o cuál sea el mecanismo más eficaz para fomentar y reclutar jóvenes investigadores que garanticen la calidad del campo en el futuro. En la actualidad el modelo lineal se ha suavizado (con la inclusión, por ejemplo, de nociones como la sostenibilidad), pero continúa existiendo una concepción bastante paralela con respecto al papel del conocimiento como motor del crecimiento social y económico. De hecho, este parece ser el mito más resistente al cambio, pese a las críticas sobre sus limitaciones. Por otro lado, el nuevo papel otorgado a la innovación parece venir a sustituir, sin modificar más que parcialmente el modelo, la antigua confianza en la ciencia básica como garantía para el bienestar social. Así, la innovación se habría convertido en el nuevo mantra del siglo XXI al que se apela, ingenua y acríticamente, como una suerte de piedra filosofal para la recuperación económica.

Además, desde la política científica continúa sin atenderse adecuadamente a la existencia de incertidumbre en la ciencia y tecnología contemporáneas, tornando complejo dar cabida a nuevos espacios de interacción con la sociedad que, aunque son reconocidos "sobre el papel", no se desarrollan de manera fructífera. Por lo que respecta a la rendición de cuentas, existen importantes matices con respecto al modelo anterior, como la elección de líneas prioritarias de investigación desde el entorno político con el fin de tratar de satisfacer las demandas sociales. Sin embargo, el

peso de la evaluación aún es en gran medida endogámico y, como consecuencia, el sistema de gestión continúa siendo fundamentalmente tecnocrático.

Finalmente, en el caso concreto de España, donde uno de los problemas endémicos ha sido la limitada inversión en ciencia y tecnología y su condicionamiento a los avatares económicos, puede considerarse que, pese a sus esfuerzos en numerosos campos, quizá la ley de 2011 ha perdido una oportunidad de oro para convertir la política científica en una verdadera "política de Estado".

REFERENCIAS

Bush, V. 1945, *Science, the endless frontier* (http://www.nsf.gov/od/lpa/nsf50/vbush1945.htm).

Coca, J.R. 2008, 'Entre ciencia, tecnología, política y sociedad anda el juego. Entrevista a Miguel Ángel Quintanilla', *ArtefaCToS* 1, 115-18.

Elzinga, A. y Jamison, A. 1995, 'Changing Policy Agendas in Science and Technology', en Jasanoff, S. *et al.* (eds.), *Handbook of Science and Technology Studies,* Sage Publications, 572-97.

García Arroyo, A. *et al.* 2007, 'Legislar sobre política científica para el siglo XXI en España', *Arbor* 727, 634-54.

Gobierno de España 1986, *Ley 13/1986, de 14 de abril, de Fomento y Coordinación General de la Investigación Científica y Técnica* (http://www.boe.es/aeboe/consultas/bases_datos/doc.php?id=BOE-A-1986-9479).

Gobierno de España 2011, *Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación* (http://www.boe.es/boe/dias/2011/06/02/pdfs/BOE-A-2011-9617.pdf).

Lafuente, A. 2006, 'Ciencia 2.0', *Mi+D Revista Electrónica* (http://www.madrimasd.org/revista/revistaespecial1/articulos/lafuente.asp).

López, S. y Santesmases, M.J. 2006, 'La ciencia en España' en González Enciso, A. y Matés Barco, J.M. (coords.), *Historia económica de España,* Barcelona, Ariel, 891-918.

MICINN 2008, 'Comparecencia de la Ministra', *Diario de Sesiones del Congreso de los Diputados* 42 (http://www.madrimasd.org/empleo/documentos/doc/Comparecencia_Ministra_Ciencia_Innovacion_Congreso_160608.pdf)

Muñoz, E. 1998, 'The Spanish System of Research: Research and Innovation in Spain', en Laredo, P. y Mustar, P. (comps.), *Research and Innovation Policies in the New Global Economy,* Cheltenham, Edward Elgar, 359-97.

Muñoz, E. y Sebastián, J. 2008, 'Exploración de la política científica en España', en Romero de Pablos, A. y Santesmases, M.J. (coords.), *Cien años de política científica en España*, Madrid, Fundación BBVA, 357-84.

Pérez Sedeño, E. y Alcalá, P. 2006, 'La Ley de la Ciencia veinte años después: ¿dónde estaban las mujeres?", *Mi+D Revista Electrónica* (http://www.madrimasd.org/revista/revistaespecial1/articulos/perezalcala.asp).

Quintanilla, M.A. 2007, 'La investigación en la sociedad del conocimiento' *Revista CTS* 8, Madrid, OEI, 183-94.

Sanmartín, J. 1990, *Tecnología y futuro humano,* Barcelona, Anthropos.

Sanz Merino, N. 2005, 'El primer contrato social para la ciencia y la posibilidad de su reforma en el siglo XXI', en Velarde Lombraña, J. *et al.*, *Studia Philosophica IV,* Oviedo, Universidad de Oviedo, 243-60.

Sarewitz, D. 1996, *Frontiers of Illusion*, Philadelphia, Temple University Press.

UNESCO 1999, *Declaración sobre la ciencia y el uso del saber científico* (http://www.unesco.org/science/wcs/esp/declaracion_s.htm)

APRENDIENDO DE LA IGNORANCIA LA AGNOTOLOGÍA COMO NUEVO ENFOQUE EN FILOSOFÍA DE LA CIENCIA

MANUELA FERNÁNDEZ PINTO
Estudiante de doctorado en Historia y Filosofía de la Ciencia
University of Notre Dame
mfernan3@nd.edu

ABSTRACT: Recent agnotological studies in the history of science have uncovered the mechanisms of ignorance construction that pervade science today. In the paper, I present agnotology as an innovative perspective that can contribute in important ways to a better understanding of scientific knowledge in philosophy of science, particularly in social epistemology of science. After examining the agnotological case of the tobacco strategy, I conclude that agnotology can help identifying the epistemological flaws and normative gaps of current scientific practice, showing the importance of a philosophical approach in science studies.
KEYWORDS: agnotology, normativity, social epistemology of science, social construction of ignorance.

Introducción

El objetivo principal de esta ponencia es presentar la agnotología, o el estudio de la ignorancia, como un enfoque metodológico alternativo en el estudio del conocimiento científico en filosofía de la ciencia. En particular, defenderé que este nuevo enfoque contribuye de forma novedosa a identificar los fallos epistemológicas y los vacíos normativos del quehacer científico actual, evidenciando la importancia de una aproximación filosófica en el marco de los estudios de la ciencia. La ponencia se dividirá en tres partes.

La primera parte describirá el marco teórico en el que deseo ubicar la ponencia. Dado que uno de los objetivos centrales de la agnotología es reconocer las dimensiones sociales de la ignorancia, mi tesis en la ponencia se limitará a ver la agnotología como una contribución al tipo de filosofía de la ciencia que tiene una visión social del quehacer científico. A este tipo de filosofía de la ciencia lo llamo *epistemología social de la ciencia.* En la segunda parte, haré una breve introducción a la agnotología y luego presentaré el caso particular de la "estrategia del tabaco", examinado por los historiadores Naomi Oreskes y Erik Conway en su libro *Merchants of Doubt* (2010). Esto con el objetivo de mostrar cuáles son las características de los estudios agnotológicos y qué tipo de conclusiones alcanzan. La tercera parte se enfocará propiamente en defender la tesis

central de la ponencia. A partir del caso examinado en la sección anterior, argumento que los estudios recientes en historia de la ciencia que han adoptado una perspectiva agnotológica muestran la necesidad de una postura normativa con respecto a la práctica científica.

1. Epistemología Social de la Ciencia

Si bien considero que la agnotología puede utilizarse como un enfoque metodológico innovador en filosofía de la ciencia, es importante aclarar que no pretendo defender esta tesis para la filosofía de la ciencia en general, sino limitarla al caso de la filosofía de la ciencia que estudia el conocimiento científico como un fenómeno social y que, por ende, toma distancia de la perspectiva filosófica tradicional que enfrenta el problema del conocimiento partiendo del sujeto individual aislado. De esta manera, en esta ponencia parto del marco de la epistemología social (de la ciencia).

La epistemología social surgió con el ánimo de responder a los retos del constructivismo de los estudios sociales de la ciencia (STS, por sus siglas en inglés). Para ello, era necesario cambiar la perspectiva individualista de la epistemología clásica, la cual está interesada principalmente en el individuo y como éste obtiene creencias verdaderas. Al concentrarse en el individuo, la epistemología tradicional no puede dar cuenta del conocimiento como un fenómeno social. La epistemología social tiene como meta incorporar los aspectos sociales de la ciencia, mientras sostiene un rechazo al relativismo, e intenta establecer una postura normativa del conocimiento.

En particular, varios filósofos trabajando en epistemología social se han dedicado principalmente al estudio del conocimiento científico (Douglas 2009; Kitcher 2001, 2011; Longino 1990, 2002; Solomon 2001). Aunque las razones para enfocarse en el conocimiento científico varían, estos filósofos parecen compartir la idea de que entender los mecanismos que llevan al conocimiento científico permitirá alcanzar un mejor entendimiento del conocimiento en términos generales. En palabras de Helen Longino, "Los argumentos sobre la naturaleza del conocimiento científico son relevantes para entender el carácter y la posibilidad del conocimiento empírico ordinario" (2002, 3).[1]

Dado que uno de los objetivos centrales de la agnotología, como veremos en un momento, es reconocer las dimensiones sociales de la ignorancia, mi tesis en la ponencia se limitará a ver la agnotología como una contribución al tipo de filosofía de la ciencia que tiene una visión social del quehacer científico, es decir, como una contribución a la epistemología social de la ciencia, a la cual le atribuyo la idea de incorporar los elementos sociales propios de la práctica científica (valores políticos y sociales, políticas públicas, instituciones educativas, comercialización y privatización, entre otros), reteniendo al mismo tiempo los objetivos normativos de la filosofía de la ciencia (e.g.,

[1] Las traducciones al español son de mi autoría.

cuáles son los límites del conocimiento científico, o la diferenciación entre lo que constituye conocimiento científico adecuado y lo que no).

2. La Agnotología y sus Resultados en Historia de la Ciencia

El historiador de la ciencia Robert Proctor ha enfocado su carrera al estudio de la construcción social de la ignorancia y, en particular, al caso de la industria tabacalera. En 2008, Proctor y Londa Shiebinger publicaron el libro *Agnotology: The Making and Unmaking of Ignorance* donde acuñaron el término *agnotology*, o *agnotología* en español, para referirse a este nuevo tipo de estudios centrados en la ignorancia. En los últimos años, la agnotología se ha visto reivindicada por importantes estudios en la historia de la ciencia (Galison 2008, Oreskes & Conway 2010; Proctor 1995, 2011).

En la introducción a *Agnotology*, Proctor desafía la concepción tradicional de la ignorancia como ausencia o vacío natural, y defiende en cambio una concepción compleja de la ignorancia, teniendo en cuenta sus dimensiones sociales:

> Es necesario pensar en la producción consciente, inconsciente y estructurada de la ignorancia; sus diversas causas y maneras de conformarse; si éstas son producidas por negligencia, olvido, miopía, extinción, secreto, o supresión. El punto es cuestionar la "naturalidad" de la ignorancia, sus causas y su distribución. (Proctor 2008, 3)

En consecuencia, Proctor distingue tres tipos de ignorancia. En primera instancia, la ignorancia es entendida como un *estado natural* u *originario*, i.e., como un vacío que debe llenarse con conocimiento. Proctor sugiere que además de esta concepción tradicional, la ignorancia debe entenderse también como un constructo social. En un segundo sentido, la ignorancia puede ser un constructo passivo o *terreno perdido.* Éste es el tipo de ignorancia que surge de las condiciones sociales en las que se hace la ciencia, inluyendo su geografía política (quiénes son los científicos y por qué, o por qué la ciencia se lleva a cabo en determinados lugares). En este sentido, la ignorancia emerge de la combinación de decisiones, circunstancias y accidentes que rodean las prácticas de construcción de conocimiento. Finalmente, la ignorancia también puede ser un constructo activo o *maniobra estratégica*: puede crearse y manternerse deliberadamente. En este sentido, la ignorancia se convierte en un producto manufacturado (2008, 4-8). Veamos más de cerca uno de los estudios clave en historia de la ciencia sobre la construcción activa de ignorancia.

La Agnotología y la Estrategia del Tabaco

En *Merchants of Doubt* (2010), Oreskes y Conway presentan la historia de un grupo de

científicos, Robert Jastrow (astrofísico, cabeza del Instituto Goddard para estudios espaciales), William Nierenberg (físico del Proyecto Manhattan y director de la Institución Scripps de Oceanografía), Frederick Seitz (físico distinguido del Proyecto Manhattan, quien estudió física del estado sólido con Eugene Wigner, luego presidente de la Academia Nacional de las Ciencias de E.E.U.U.), y Fred Singer (científico especializado en cohetes, primer director del Servicio Nacional de Clima por Satélite), quienes estuvieron seriamente inmiscuidos en la construcción deliberada de ignorancia. Por medio de la manipulación y la supresión de información, el uso inapropiado de sus credenciales científicas, apelaciones inadecuadas al balance periodístico y ataques personales a otros científicos, este grupo de físicos renombrados logró engañar al público y promover la ignorancia por encima del consenso científico.

Uno de los ejemplos clave de Oreskes y Conway es la controversia relacionada con el tabaco y el cáncer de pulmón. El caso de las tabacaleras es paradigmático, pues a partir de éste se estableció la agenda para una serie de ataques al consenso científico desde el cigarrillo y la capa de ozono, hasta el caso del calentamiento global.

Oreskes y Conway explican cómo los presidentes de las mayores tabacaleras en los E.E.U.U. decidieron colaborar con la firma de relaciones públicas Hill & Knowlton, luego de que científicos del Sloan-Kettering Institute demostraran, en 1953, que el tabaco aplicado a la piel de los ratones causaba cancer. En los años siguientes, las tabacaleras junto con H&K diseñaron una estrategia para oponerse a los descubrimientos científicos contra el uso de cigarrillos (2010, 14-16).

Si bien Orekes y Conway no la caracterizan de esta manera, es posible distinguir al menos cinco elementos clave en la estrategia. Primero se presenta un ataque personal a científicos que llevan a cabo investigaciones legítimas. Por ejemplo, cuando la EPA publicó el reporte titulado *Respiratory Health Effects on Passive Smoking* (1992), en el que se atribuía enfermedad crónica y muerte a los fumadores pasivos, Fred Singer produjo otro reporte acusando a la EPA de hacer "junk science", y de sostener posiciones que no estaban respaldadas por la evidencia, con el objetivo de expandir el control gubernamental (2010, 140-2). El informe de Singer estaba financiado por una beca del Tobacco Institute, a través de un think tank, la Institución Alexis de Toqueville.

Un segundo punto clave de la estrategia fue generar incertidumbre con respecto a la evidencia disponible. Como sugieren los memos de H&K dirigidos a las tabacaleras, la idea principal era enfatizar cada controversia y duda que pudiera encontrarse en las investigaciones. Así, el objetivo no era negar la conexión entre el cigarrillo y el cáncer de pulmón, sino crear dudas sobre las investigaciones que obtenían esa conclusión. Como lo estableció un famoso memorandum interno de Brown and Williamson en 1969: "La duda es nuestro producto, ya que es el mejor medio para competir con el "conjunto de hechos" que existe en la mente del público" (2010, 34).

En tercera instancia, la industria tabacalera fortaleció su estrategia, con la reestructuración del Consejo para la Investigación del Tabaco (Council for Tobacco Research). El consejo fue fundado por las tabacaleras para cuidar su propia imagen, mediante la vigilancia constante de sus relaciones públicas, a partir de una relación cercana con H&K. El CTR se convirtió en el centro de operaciones mediante el cual las diferentes firmas coordinaban sus investigaciones y vigilaban las campañas anti-tabaco.

La industria tabacalera también inició una estrategia de creación y fortalecimiento de conexiones con el sistema académico, a partir de la financiación de proyectos de investigación médica en universidades, la creación de becas para financiar estudiantes de medicina, la conexión con miembros de las asociaciones más relevantes en estos temas de salud, y la publicación y distribución masiva de panfletos, artículos e informes que apoyaran su causa.

Por último, toda la estrategia se vio legitimada por científicos prominentes, entre los que se destacan Jastrow, Nierenberg, Seitz y Singer, quienes utilizaron sus credenciales científicas para desafiar el consenso científico y promover incertidumbre.

Esta estrategia resultó ser supremamente exitosa: la industria no perdió ningún caso en la corte hasta los años 90 y la FDA no pudo regular el tabaco como sustancia adictiva hasta el 2009 (2010, 33). Adicionalmente, la "estrategia del tabaco" también fue usada como oposición al consenso en otros debates científicos como el debate sobre la capa de ozono, la lluvia ácida y el calentamiento global. El hecho de que la misma estrategia haya sido aplicada exitosamente para desafiar resultados científicos en diferentes campos hace de la "estrategia del tabaco" un punto crucial de la historia de la ciencia y su política en las últimas décadas. En particular, ésta puede considerarse una estrategia de construcción activa de ignorancia, o *maniobra estratégica*: su principal objetivo fue promover deliberadamente la ignorancia mientras ocultaba conocimiento.

3. La Agnotología Como Nuevo Enfoque en Filosofía de la Ciencia

La agnotología puede considerarse un enfoque innovador en filosofía de la ciencia, pues tiene el potencial de contribuir al análisis filosófico del conocimiento científico. Primero, los estudios agnotológicos que han surgido en historia de la ciencia muestran las fallas epistemológicas de las prácticas científicas de la actualidad (entre las que podemos destacar los problemas que representa la financiación de estudios médicos, estudiantes, revistas especializadas, etc. por parte de la industria tabacalera, las farmacéuticas, las petroleras, etc.). Así, la agnotología hace evidentes los mecanismos de construcción de ignorancia que evitan la producción, divulgación y aceptación del conocimiento científico, abriendo la posibilidad de un examen epistemológico sobre estas prácticas. Bajo la perspectiva de la agnotología podemos ver de forma más clara cuáles son los obstáculos

reales que el conocimiento científico enfrenta hoy, cuando la ignorancia es construida socialmente.

Además, asumir el enfoque agnotológico permite evidenciar los fallos normativos de las prácticas científicas de la actualidad y, de esta manera, muestra a los filósofos una forma innovadora de contribuir a los estudios de la ciencia. Por ejemplo, Oreskes y Conway parecen defender una visión del conocimiento científico que requiere compromisos normativos. Se muestran confiados en la habilidad de los científicos para diferenciar entre lo que constituye ciencia aceptable o no: "Los científicos se muestran confiados en poder identificar mala ciencia cuando la ven. Es ciencia obviamente fraudulenta, en la que los datos han sido inventados, mezclados, manipulados." (2010, 153); y confían en los expertos para determinar el mejor curso: "En materia de ciencia debemos confiar en nuestros expertos, pues si bien la ciencia moderna no nos da certeza, sí tiene un récord sólido." (2010, 272-3). *Merchants of Doubt* puede considerarse en parte una denuncia pública de lo que constituye "mala ciencia", o "malos científicos". Oreskes y Conway sugieren, si bien no explícitamente, una normatividad científica que podría en principio oponerse a la construcción de ignorancia en casos como el del calentamiento global o el tabaco.

La controversia que rodea los casos del calentamiento global y el tabaco requieren una evaluación no sólo legal y moral, sino también epistemológica. En estos casos, querríamos decir (como sugieren Oreskes y Conway) que estos científicos actuaron en detrimento del conocimiento, que impidieron el curso normal de la investigación científica, y que ocultaron la verdad al público. Pero para poder juzgarlos en términos epistemológicos necesitamos una postura normativa de la ciencia que nos permita decir que lo que estos científicos hicieron no fue ciencia en un sentido apropiado de la palabra, que lo que estaban fomentando no era conocimiento y que debido a sus acciones se incrementó la ignorancia. Necesitamos una visión del conocimiento científico que capture tanto los componentes sociales de los ejemplos (las posturas políticas, las conexiones entre la industria y el gobierno, la reputación de los científicos, la experiencia profesional, etc.) como también la normatividad requerida para evaluar entre prácticas científicas aceptables e inaceptables.

La filosofía de la ciencia, dedicada tradicionalmente al estudio del pensamiento científico independientemente de sus implicaciones sociales, apenas comienza a entender y a responder a los retos normativos de la agnotología (e.g., Lacey 2005). Dados los problemas epistemológicos que estos estudios han identificado y dada la importancia de una normatividad que permita afrontar estos problemas, considero que la agnotología es un enfoque en cierta medida privilegiado para la filosofía de la ciencia. Al concentrarse en los mecanismos de obstrucción, la agnotología permite visibilizar fallos en la práctica científica que han permanecido ocultas en los enfoques tradicionales. Cómo evaluar estas fallas (quizá con el ánimo de prevenirlas) o cómo explicar por qué éstas constituyen problemas epistemológicos serios son preguntas clave, a cuya respuesta la filosofía de

la ciencia tiene aún mucho qué contribuir.

En resumen, la agnotología se presenta como un nuevo enfoque que permitirá a la filosofía identificar los problemas que la construcción de ignorancia trae para el correcto desarrollo e implementación de la ciencia. El reto de la agnotología es un reto normativo: se trata de asumir las dimensiones sociales de la ignorancia (no sólo del conocimiento) y proveer una postura normativa de la ciencia en la que se tenga en cuenta este aspecto, es decir, que sea capaz de identificar y rechazar las prácticas de producción de ignorancia que son epistemológicamente dañinas. Así, el enfoque agnotológico presiona a la filosofía a desarrollar una postura más sofisticada sobre la ciencia que mantenga las metas normativas de la filosofía, pero que al mismo tiempo sea más relevante y apropiada para evaluar la ciencia de hoy. Después de más de cuatro décadas de defender la normatividad en contra de los retos historicistas y constructivistas, la agnotología ofrece a la filosofía de la ciencia una nueva oportunidad de contribuir de manera significativa a los estudios de las ciencia.

REFERENCIAS

Douglas, H. 2009, *Science, Policy, and the Value-Free Ideal*, Pittsburgh, University of Pittsburgh Press.

Galison, P. 2008, 'Removing Knowledge: The Logic of Modern Censorship', in Proctor, R. & Schiebinger, L. (eds.) *Agnotology: The Making and Unmaking of Ignorance,* Stanford, Calif, Stanford University Press, 37-54.

Kitcher, P. 2001. *Science, Truth, and Democracy*. New York, Oxford University Press.

———. 2011. *Science in a Democratic Society*. Amherst, NY, Prometheus Books .

Lacey, H. 2005, *Values and Objectivity in Science and Current Controversy About Transgenic Crops*. Lanham, MD, Lexington Books.

Longino, H. 1990, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ, Princeton University Press.

———. 2002, *The Fate of* Knowledge. Princeton, NJ, Princeton University Press.

Oreskes, N., & Conway, E. 2010, *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. New York, Bloomsbury Press.

Proctor, R. 2011, *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition*. Berkeley, University of California Press.

———. 1995, *Cancer Wars: How Politics Shapes What We Know and Don't Know about Cancer*. New York, Basic Books.

Proctor, R., & Schiebinger, L. 2008, *Agnotology: The Making and Unmaking of Ignorance*. Stanford, Calif, Stanford University Press.

Solomon, M. 2001, *Social Empiricism*. Cambridge, Mass, MIT Press.

CULTURA CIENTÍFICA Y MUSEOS DE CIENCIA: UNA HISTORIA PARALELA[1]

BELÉN LASPRA PÉREZ
Departamento de filosofía, Universidad de Oviedo

ABSTRACT: The history of science and technology museums and the evolution of the scientific literacy concept are two stories with enough common ground to allow a reconstruction parallel development. It is possible to detect links between the way that science and technology museums conceive the relationship between science and public and the paradigms that have guided the research of scientific culture, characterized by a deficit model.
KEY WORDS: Scientific literacy, museum of science and technology, deficit model.

1. Primera generación de museos – ¿Pre modelo de déficit?

La historia de los museos de ciencia y tecnología y la evolución del concepto de cultura científica son dos historias con suficientes puntos en común como para permitir una reconstrucción en paralelo de su desarrollo. Es posible detectar nexos entre el modo en que los museos de ciencia y tecnología conciben la relación entre la ciencia y el público y los paradigmas que han guiado la investigación de la cultura científica, caracterizados por el modelo de déficit.

Los primeros museos de ciencia y tecnología son concebidos como reservorios de objetos de colección, símbolo del patrimonio científico y tecnológico de la civilización occidental. Se limitan a mostrar la ciencia como una serie de catálogo de realizaciones técnicas de un país o de una época determinada (Pérez, *et al* 1998). Eran habitaciones llenas de objetos organizados para mostrar las maravillas de la ciencia y la tecnología de la época al público (Baird 1986) compuesto por estudiantes y en general por la clase alta.

1
 Este trabajo ha sido realizado con el apoyo del proyecto «Políticas de la cultura científica (FFI2011-24582) y del Programa Severo Ochoa de Becas Predoctorales del Principado de Asturias.

A eso se reduce en ellos la educación, a colocar los objetos en orden cronológico en vitrinas, y a esperar a que el visitante se enriquezca caminando entre ellas. La museología de la ciencia se basaba en los mismos principios de los que parte el método científico[2]. El modo de expresarse de los museos y el modo de exhibir sus colecciones dependía de la concepción de la ciencia, y de los que se entendía por conocimiento científico que precisamente era el establecido por la comunidad científica y controlado por ella (Schiele 2008).

Son ejemplos de estos primeros museos el Conservatorie National des Arts et Métiers de París y el South Kenshington Museum de Londres. El Conservatoire, fundado en 1794, fue creado para "guardar prototipos de máquinas, modelos, útiles, diseños y libros de relativos a todo género de artes y oficios […] que daban una idea de la evolución de las técnicas del siglo XVIII" (Martín 1964, 112). En él la didáctica de las técnicas se realizaba mediante la presentación comentada de los objetos y la demostración, llevada a cabo por especialistas, de cómo estos objetos se encendían y se apagaban (Hernández, 1998). El South Kenshington, actualmente conocido como London Science Museum, se abrió al público en 1857 como consecuencia de la Exposición Universal de Londres en 1851. La popularidad que alcanzó el Cristal Palace, lugar donde se celebró la Exposición, llevó al Príncipe Albert a fundar varios establecimientos de carácter educativo, entre ellos el South Kenshington (Burton 1999), destinado a albergar objetos y colecciones que testimoniaban los avances realizados a lo largo de la historia de la ciencia, la tecnología y la medicina.

Estos primeros museos de ciencia y tecnología museos eran, en definitiva, reservorios de objetos y no se puede afirmar que respondieran a una necesidad de promoción de la cultura científica con el objetivo de formar ciudadanos capaces de tomar decisiones en la arena pública. Albergaban las evidencias materiales de las leyes de la naturaleza y del lugar del hombre, que preservaban y mostraban para la ilustración del público (Schiele 2008) así que en cierto modo constituyen un estadio previo al modelo de déficit cognitivo, ya que hay en ellos una preocupación por la alfabetización científica.

2

 La realidad existe independientemente de la representación que nos hacemos del mundo; la realidad es independiente de cualquier lenguaje usado para describirla, la verdad es representación de la realidad, y el conocimiento es objetivo (Searle 1995).

Pero éstos museos acabaron quedando obsoletos, dando paso a una segunda generación de museos que responde ya a la necesidad de incrementar la cultura científica de la sociedad, aunque siguen reducíendo el concepto de cultura científica al de alfabetización científica caracterizado por Bauer (ver más adelante).

2. Segunda generación de museos - Paradigma 'Scientific Literacy'

La segunda generación[3] de museos se caracteriza por añadir a la conservación y presentación del patrimonio científico una decidida voluntad pedagógica y cultural (Pérez, *et al* 1998). Esta voluntad pedagógica se traduce en una reducción de los artefactos que se muestran[4] y en potenciar las demostraciones científicas (Danilov 1973). El público se contempla como una masa potencialmente alfabetizable en materias de ciencia y tecnología, pero su consideración no irá más allá de una mera entidad pasiva que recibe la información de la comunidad científica. En general los museos de esta generación presentan las posibilidades de la ciencia y la tecnología como solución a los problemas de la humanidad (Baird 1986).

Son ejemplos de esta generación el Deutschest Museum für Wissenschaft und Technik de Múnich y el Palais de la Dècouverte de París. Inspirándose en la Exposición Internacional de la Electricidad en 1881, celebrada en París, Oskar von Miller fundó el Deutches Museum en 1903 con el objetivo de "crear una institución educativa para todos en la que se impartieran los fundamentos de la ciencia la tecnología mediante el encuentro con sus obras maestras" (Knerr 2000, 8). El Palais, construido para albergar la Exhibición Internacional de 1937, se transformó en un establecimiento permanente en 1938. La pretensión del museo era que el público comprendiera la importancia de los descubrimientos científicos en la creación de la civilización (Schiele 2008), y su objetivo declarado era la difusión de la cultura científica (Hernández 1998).

3

 Distintos autores señalan diferentes etapas en la historia de los museos de ciencia (Baird 1986, Danilov 1976, Pérez, *et al* 1998). No se pretende aquí usar el término en sentido estricto, tan sólo para marcar las diferencias entre unas etapas y otras.

4

 Por primera vez los fondos del museo y los contenidos de la exposición comienzan a separarse, siguiendo la estela abierta por la forma de organizar la exhibición de la Exposición Internacional de Chicago en 1933, que clasificaba los objetos temáticamente.

Suele datarse el inicio del primer paradigma de la investigación de la cultura científica en 1960. La primera encuesta sobre la percepción social de la ciencia fue dirigida por Davis en 1957. Los resultados del estudio mostraron el escaso conocimiento, de la sociedad estadounidense, sobre materias de ciencia y tecnología y un bajo apoyo en general, y prescribía que la falta de apoyo a la ciencia era debido a un déficit de conocimientos científicos[5]. Si el diagnóstico de Davis era correcto, la solución era "educar a los ciudadanos, elevando su nivel de formación científica, [lo que] conduciría a mejorar su apreciación y a disminuir su apatía o reticencias" (Cortassa 2010, 52) en orden a conseguir ciudadanos capaces de ejercer su derecho a participar en asuntos de política pública. La alfabetización científica consistía pues en "conocer los hechos de libros de textos básicos de la ciencia, comprender métodos tales como el razonamiento de probabilidad y el diseño experimental, valorar positivamente los resultados de la ciencia y la tecnología, así como el rechazo de creencias supersticiosas" (Bauer 2009, 223).

Los museos de ciencia y tecnología, como instituciones que estaban comenzando a incluir abiertamente entre sus fines la educación son lugares propicios para el fomento de la cultura científica. Inconscientes de las acciones políticas[6] que se tomarían a raíz de los resultados de la encuesta de Davis, estos museos ya percibían la alfabetización científica "como una condición necesaria para cimentar el compromiso de la sociedad civil" (Cortassa 2010, 164)

3. Tercera generación - Paradigma 'Public Understanding of Science'

Los primeros años del siglo XX fueron una época de bonanza para la ciencia[7], pero tras la Segunda Guerra Mundial, las ventajas y desventajas del avance científico-

5

En realidad se pasó de una hipótesis correlacional: "existe una correlación entre el nivel de conocimientos científicos y el grado de interés por la ciencia" a una hipótesis causal: "el grado de interés por la ciencia depende del nivel de conocimientos científicos". (Véase Cortassa 2010)

6

En Norte América se centraron en mejorar la educación de los ciudadanos, mientras que en Inglaterra la Royal Society asumió el compromiso de la expansión de la cultura científica. (Cortassa 2010)

7

Avances de indiscutible relevancia en distintos campos científicos se producen en este siglo: Albert Einstein formula la teoría de la relatividad especial, James Watson y Francis Crick descubren la

tecnológico se fueron poniendo de manifiesto, especialmente a partir de 1960 y tras una serie de desastres artificiales. La sociedad demandaba una perspectiva de las cuestiones y riesgos asociados a la ciencia y a la tecnología, y los museos comenzaron a incluir en sus exposiciones aspectos sociales y cuestiones relacionadas tanto con los beneficios de la ciencia como con los perjuicios (Schiele 2008), convirtiéndose así en mediadores entre la comunidad científica y los legos. Es en esta década de los 60 cuando se produce una explosión de museos de ciencia por todo el mundo, muchos bajo el rótulo de 'centro de ciencia'. Y es también esta tercera generación de museos la que enlaza directamente con el segundo paradigma de la investigación en cultura científica.

El segundo paradigma de la cultura científica comienza en 1985, da la fecha la publicación del Informe de la Royal Society de Londres, dirigido por Boodmer, en ese mismo año. Los datos del estudio pusieron de manifiesto el escaso apoyo a la ciencia por parte de la sociedad, y concluía que esa falta de apoyo era debida a una falta de interés por parte de la sociedad en temas relacionados con la ciencia y tecnología (Bauer 2009). Si el paradigma de 'Scientific Literacy' era caracterizado por un déficit de conocimiento, el paradigma de 'Public Understanding of Science' se caracteriza por un déficit de interés. El diagnóstico fue la falta de interés, y el tratamiento se llevaría a cabo mediante campañas de comunicación y divulgación del conocimiento científico.

En la década de los 70, centros de ciencia como el Exploratorium de San Francisco o el Ontario Science Center de Canadá, ambos inaugurados en 1960 y ambos de referencia internacional, surgen pensados y diseñados bajo principios que responden abiertamente a la comunicación[8]. La idea es que una buena comunicación de la ciencia incrementa el interés por la ciencia y ayuda a mejorar la alfabetización científica, exactamente el mismo tratamiento que prescribía la Royal Society. Este giro hacia la comunicación es la base de lo que se conoce actualmente como el movimiento *hands-on* de los centros de ciencia, cuyo principal objetivo es superar las limitaciones de los

estructura del ADN, se completa la tabla periódica de los elementos, se formulan las primeras explicaciones científicas sobre el origen del universo, entre otros.

8

No todos estaban dispuestos a adoptar el principio comunicativo, como reflejan las tensiones que rodearon la creación del Smithsonian. Su primer secretario, Josep Henry, rechazaba cualquier actividad que no fuera la de la investigación, aunque finalmente no pudo impedir el desarrollo de lo que sería el Smithsonian's National Museum of American History de Washington, abierto al público en 1964 (Schiele 2008).

museos de ciencia y tecnología tradicionales enfatizando el carácter educativo basándose en métodos que propician la interacción y la participación (Pérez *et al.* 1998).

4. Cuarta generación - Paradigma 'Science in and of Society'

El tercer paradigma de investigación de la cultura científica comienza en 1990. El informe *Science and Society (House of Lords)* publicado en 2000 fijaba la falta de apoyo a la ciencia por parte de la sociedad en un déficit de confianza (Bauer, 2009). La participación social será la estrategia propuesta ante el clima de desconfianza hacia la ciencia.

En el caso de los centros de ciencia, la participación ha seguido una evolución similar. En los años 60, la participación del visitante en las exhibiciones de un centro de ciencia consistía en apretar botones. Esta forma de entender la participación, demasiado simple, llevó al descrédito de aquellos centros que la practicaban. Entre 1980 y 1990, y a la luz de los avances acontecidos en el campo de la didáctica de las ciencias, comenzaron a surgir dudas sobre la eficacia de la metodología interactiva. Estos estudios destacaban que la manipulación resultaba insuficiente para generar aprendizaje, y que el contexto teórico desde el que el visitante accede al museo, su forma de ver el mundo y su escala de valores han de ser tenidos en cuenta (Guisasola, *et al* 2000). Por primera vez comienza a hablarse de públicos. Los museos de esta cuarta generación cuya función es la de comunican la ciencia (Pérez, *et al* 1998) y que incluye específicamente a los centros de ciencia, comienza a entenderse la participación como "una asociación compleja entre los visitantes y la exposición, que debe ofrecer perspectivas para la manipulación, la experimentación, y la estimulación de los sentidos cuanto sea posible" (Orchiston, *et al* 1984, 37). Sin embargo en la práctica se reduce a una mera sofisticación de la interfaz, sustituyendo la vitrina por el botón.

En este sentido, algunos autores afirman que pervive en los museos de ciencia y tecnología, así como en los centros de ciencia, una concepción anacrónica de ciencia, al margen de problemas y controversias (Delicado 2009), tienden a presentar la ciencia fragmentada y descontextualizada, sin preocuparse por transmitir cualquier tipo de comprensión global y sistemática (Durant 2001) lo que dificulta el tratamiento de temas científicos que involucran factores sociales, políticos o económicos. Así que las

impresiones de Champagne (1975) en su visita a uno de los primeros centros de ciencia de la historia, sigue siendo válido: cualquier referencia a la dimensión ética de la ciencia y la tecnología, cualquier pregunta acerca de las consecuencias de sus productos y procesos, así como cualquier vinculación con los problemas importantes del mundo o la vida diaria es ignorada (Champagne 1975).

5. Un nuevo tipo de museo

Es necesaria la creación de un espacio que permita la confluencia de la ciencia, la tecnología y la sociedad. A esta especie nueva de museo será denominada "museo de ciencia, tecnología y sociedad"[9].

En este nuevo tipo de museos la ciencia debe estar contextualizada en lo social: mostrando los resultados de la ciencia, el método empleado para obtenerlos y abordando temas de controversia, involucrando a los visitantes. La comunicación debe ser entendida como un proceso no lineal, esto implica que puede originarse tanto en contextos expertos como en contextos no expertos o, al menos, cuestionar la tradicional distinción entre expertos y legos. La recepción de la comunicación debe abordarse como un conjunto complejo de procesos formativos a los que los diferentes públicos, como entidad activa, acceden desde sus contextos particulares.

Es decir, un museo de ciencia, tecnología y sociedad debe ser sensible al contexto particular desde el cual el visitante se acerca a la ciencia, así como a sus conocimientos previos, sus actitudes y sus valores, entendiendo que el visitante tiene derecho a rehacer su verdad por sí mismo, evitando enviar mensajes blindados por la tradición de la autoridad científica (Wagensberg 2000), y propiciando que el ciudadano entienda el proceso entre el descubrimiento científico y la aplicación tecnológica, así como el debate de las implicaciones sociales, morales y económicas de las mismas, además de favorecer un debate interdisciplinar que permita explicitar las diferentes facetas de la ciencia, su filosofía, sociología e historia (Guisasola, *et al* 2000).

9

La propuesta de un Museo de Ciencia, Tecnología y Sociedad se presentó en el artículo García, M. y Laspra, B. (2011) 'Cultura científica en los museos de ciencia y tecnología' en Díaz, I. y Muñoz, A. (eds.) Participación y cultura científica en contexto internacional, Madrid, Catarata.

REFERENCIAS

Baird, D.M. 1986, 'Science museums in the modern world', *Curator*, 29/3, 213-220.

Bauer, M.W. 2009, 'The evolution of Public Understanding of Science-Discourse and comparative evidence', *Science, Technology and Society*, 14/2, 221-240.

Burton, A. 1999. *Vision and accident: the story of the Victoria and Albert Museum*, Londres, V&A Publications.

Cortassa, C.G. 2010, 'Del déficit al diálogo, ¿y después? Una reconstrucción crítica de los estudios de comprensión pública de la ciencia', *Revista CTS*, 15/5, 47-72.

Champagne, D.W. 1975 'The Ontario Science Center in Toronto: Some impresions and some questions', in *Educational Technology, 15/*8, 36-39.

Danilov, V. 1976, 'America's contemporary science museums', *Museum Journal*, 76/75, 145-147.

Delicado, A. 2009, 'Scientific controversies in museums: notes from a semi-peripherical country', in *Public Understanding of Sciencie , 18/*6, 759-767.

Durant, J. 2001, 'Introduction', in J. Durant (ed.) *Museums and the public understanding of science*, London, Science Museum, 7-11.

Guisasola, J. e Intxausti, S. 2000, 'Museos de ciencia y educación científica: una perspectiva histórica', en *Alambique*, 26, 7-14.

Hernández Hernández, F. 1998, *El museo como espacio de comunicación*, Gijón, Trea.

Knerr, G. 2000, 'Museos de tecnología: nuevos públicos y nuevos socios', *Museum Internacional*, 208, 8-13.

Martín, C. S. 1964, *Hacia un museo español de la ciencia y de la técnica. Informe sobre los museos de ciencia más importantes del mundo*, Madrid, Instituto Torres Quevedo.

Orchiston, W. and Bhathal, R.S, 1984, 'Introducing the Science Centrum: a new type of science museum', in *Curator*, 27, 33-47.

Pérez, C.; Díaz, M.P.; Echeverría, I.; Moretín, M. y Cuesta, M. 1998, *Centros de ciencia: Espacios interactivos para el aprendizaje*, País Vasco, Servicio editorial de la universidad del País Vasco.

Schiele, B. 2008, 'Science museums and science centers', in Bucchi, M and Trench, B. (eds.) *Handbook of public communication of science*, London-New York, Routledge, 27-39.

Searle, J.R. 1995, 'Postmodernism and the Western rationalist tradition', in Arthur, J. and Shapiro, A. (eds,), *Campus wars – Multiculturalism and the Politics of Difference*, Boulder, CO, Westview Press, 28-48.

Wagensberg, J. 2000, 'Principios fundamentales de la museología científica moderna' en *Alambique,* 26, 15-19.

LAS REGLAS DE INFERENCIA COMO HIPÓTESIS EMPÍRICAS. LA INFORMACIÓN MECANÍSTICA EN LA EVALUACIÓN DE RIESGOS[1]

JOSÉ L. LUJÁN, OLIVER TODT, JUAN BAUTISTA BENGOETXEA
Universitat de les Illes Balears

ABSTRACT: This paper focuses on the way in which mechanistic information has influenced the methodological controversies in risk assessment. Even though this kind of information serves to empirically test inference guides, it cannot close the controversy around the dose-response extrapolation models. This conclusion allows us to suggest that non-cognitive values cannot be rejected in risk assessment activities.
KEY WORDS: Mechanistic information, mechanism, risk assessment, inferential guide, methodological controversy, non-cognitive values.

En este trabajo analizamos el modo en el que la información mecanística ha influido sobre las controversias metodológicas en la evaluación de riesgos. Aunque dicha información sirve para someter a contrastación empírica las guías de inferencia, no consigue clausurar la controversia respecto a los modelos de extrapolación dosis-respuesta. Esta conclusión sugiere que los valores no cognitivos no se pueden eliminar en la evaluación de riesgos.

1. Controversias metodológicas en la evaluación de riesgos

El objetivo principal de la evaluación de riesgos es obtener información sobre el riesgo que determinadas sustancias entrañan para el entorno y la salud pública, información que posteriormente se utiliza en la toma de decisiones (elaboración de regulaciones) con el objetivo de minimizar tales riesgos. Durante la última década se han producido diferentes controversias metodológicas en el ámbito de la gestión de riesgos. Estas controversias metodológicas se han centrado en los siguientes temas: 1) carga de la prueba; 2) reglas de inferencia; y 3) estándares de prueba (Luján 2005). En este trabajo nos centraremos principalmente en las reglas de inferencia.

Las dos metodologías básicas de la evaluación de riesgos son los estudios epidemiológicos y los bioensayos, y en ambas aparecen indeterminaciones que es necesario salvar mediante extrapolaciones de los datos disponibles a situaciones reales de exposición (Cranor 1993, Shrader-Frechette 1994, López Cerezo & Luján 2000). Ante esto, las agencias reguladoras elaboran guías en

[1] Trabajo elaborado en el marco del proyecto de investigación «La explicación basada en mecanismos en la evaluación de riesgos», referencia FFI2010-20227/FISO, de la Secretaría General de Ciencia, Tecnología e Innovación del Gobierno de España, cofinanciado por fondos FEDER de la Comisión Europea.

las que explícitamente proponen reglas de inferencia en función de las sustancias investigadas. Por ejemplo, en la evaluación de productos químicos no carcinogénicos, habitualmente se utilizan modelos de extrapolación con umbrales, mientras que en los estudios sobre sustancias potencialmente carcinógenas se utilizan modelos lineales (sin umbrales). Se asume por tanto que en el caso de sustancias no carcinógenas hay una dosis por debajo de la cual no se producen efectos tóxicos, mientras que en el caso de las substancias carcinógenas, aunque las dosis sean pequeñas, cabe la posibilidad de que aquéllas causen alteraciones en el ADN que conduzcan a la aparición de tumores.

Dado que el objetivo último de la evaluación de riesgos es proporcionar información fundamental para la toma de decisiones públicas, en las controversias metodológicas a menudo se apela también a valores no cognitivos como la protección de la salud pública y del ambiente. Un modelo lineal es más protector de la salud pública que un modelo con umbrales, puesto que el primero es coherente con la hipótesis según la cual, aunque las dosis sean pequeñas, puede haber efectos negativos sobre la salud. En las discusiones que han tenido lugar los últimos años en torno a la elaboración de las nuevas directrices de la EPA (*Environmental Protection Agency*) para las evaluaciones de riesgo de sustancias carcinógenas, se habla explícitamente de modelos de extrapolación que sean tanto científicamente defendibles como protectores de la salud pública.

2. La información mecanística y las guías de inferencia

Durante la última década, una parte importante del debate metodológico en la evaluación de riesgos se ha centrado en la 'información mecanística', que puede ser relativa al *modo de acción* y/o al *mecanismo de acción* de una sustancia. El *modo* hace referencia a una secuencia de sucesos consecuencia de la interacción entre un organismo y un tóxico que conducen a la respuesta observada (por ejemplo, la mutagenicidad, la mitogénesis o la inhibición de la muerte celular). El establecimiento de los mecanismos de acción consiste precisamente en determinar la secuencia de los diferentes estadios que son clave para que aparezca una enfermedad como el cáncer. Los *mecanismos* de acción se refieren a la comprensión detallada, normalmente a nivel molecular, de los sucesos a los que se refieren los modos de acción (USEPA 2003). La diferencia entre estos dos tipos de información mecanística está relacionada con el nivel de comprensión del fenómeno. En la evaluación de riesgos se utiliza principalmente la información relativa a los modos de acción, pues conocer los mecanismos de acción es mucho más difícil (White *et al*. 2008).

La información mecanística permite someter a contrastación empírica las guías de inferencia (Clewell 2005). La idea es que la información sobre los sucesos que conducen a una enfermedad causada por una sustancia es pertinente para seleccionar el modelo que relaciona la dosis con la

respuesta. Como hemos señalado, en ocasiones puede ser muy difícil conocer si a bajas dosis una sustancia causa los mismos efectos que han sido observados a altas dosis, por lo que es necesario decidir entre diferentes modelos de extrapolación. Sin embargo, en algunos casos puede ser más fácil conocer si a bajas dosis la sustancia desencadena o no un episodio clave que forma parte de la secuencia que conduce al efecto en cuestión. También el modo de acción de una sustancia sirve para evaluar las extrapolaciones de animales a humanos (Clewell 2005).

Por lo tanto, la información mecanística puede ser importante en los debates sobre las guías de inferencia porque las dos metodologías arriba mencionadas de la evaluación de riesgos (estudios epidemiológicos y bioensayos) abordan los efectos toxicológicos de las sustancias como una caja negra: no analizan los mecanismos causales. Ahora bien, en muchos casos la información sobre la relación entre estructura química y actividad biológica, o sobre los daños genéticos y cromosómicos, puede contribuir a determinar la relación causal entre exposición y enfermedad (Cranor 2004, Cranor 2011). Por lo tanto, la información mecanística tiene la capacidad de cambiar el debate sobre los valores (cognitivos y no cognitivos) en la evaluación de riesgos.

3. Ciencia y valores en la evaluación de riesgos

Para algunos autores las guías de inferencia no son eliminables en la evaluación de riesgos, y la elección entre propuestas alternativas no está determinada por la evidencia empírica. La conclusión es que la evaluación de riesgos no es normativamente neutral, y que la elección entre reglas de inferencia alternativas se ha de realizar también sobre la base de consideraciones relativas a los costes sociales de los errores. Esta característica, junto con el hecho de que se trata de una actividad científica cuyo objetivo es orientar la toma de decisiones, ha convertido la evaluación de riesgos en un ámbito destacado para el estudio de la interacción entre valores cognitivos y no cognitivos en el desarrollo del conocimiento científico (Cranor 1997, Douglas 2006, Shrader-Frechette 2004, Wandall 2004). Veamos dos ejemplos de este tipo de análisis.

Por un lado, Shrader-Frechette (2001, y especialmente 2004) ha defendido un análisis metacientífico que muestre en evaluaciones de riesgo concretas la dependencia de las elecciones metodológicas de supuestos no siempre bien justificados científicamente. Este enfoque de Shrader-Frechette puede interpretarse del siguiente modo: la función legítima de los valores no cognitivos es influir en el análisis metacientífico. Son los valores no cognitivos los que conducen a un escrutinio crítico de la metodología utilizada en la evaluación de riesgos y, finalmente, a una mejora de dichas evaluaciones, a una mejor ciencia. Esta es una influencia legítima de los valores no cognitivos en la evaluación de riesgos.

La función general que Douglas (2006, 2009) asigna a los valores no cognitivos, por otro lado, es la de influir en la determinación de los estándares de prueba; esto es, los valores no cognitivos pueden influir en determinar el nivel de evidencia que se va a exigir en la aceptación de hipótesis en la evaluación de riesgos. Considera además que las reglas de inferencia pueden estar infradeterminadas por los valores cognitivos.[2] En uno de sus trabajos plantea la situación de elegir entre un modelo con umbral y otro sin umbral (lineal), sin que ninguno de los dos modelos posea más poder explicativo o más alcance que el otro. Douglas defiende que en casos similares los valores sociales y éticos pueden diferenciar entre los dos:

> Los valores cognitivos no diferencian entre los dos modelos. Ninguno posee más poder explicativo o alcance que el otro. Sin embargo, los valores sociales y éticos pueden diferenciar entre los dos. En una situación de incertidumbre sustancial, el modelo lineal de extrapolación es generalmente más protector de la salud pública que el modelo con umbral, mientras que el modelo con umbral generalmente protege contra la sobre-regulación. Si los datos no pueden decidir entre los dos, nuestras preocupaciones sobre la aceptabilidad de la incertidumbre, y las consecuencias de los errores, si pueden. Donde los valores cognitivos no nos iluminan, lo hacen las consideraciones sociales y éticas (Douglas 2006, 8).

Cuando las reglas de inferencia están infradeterminadas por los valores cognitivos es legítimo apelar a los valores sociales y éticos relativos a las consecuencias no cognitivas de los errores.

Conclusiones

La inclusión de la información mecanística en la evaluación de riesgos influye decisivamente en el debate sobre la función de los valores no cognitivos. Este debate deja paso a otro en el que se discute sobre las situaciones en que tales valores pueden o no ser considerados.

Este desplazamiento del debate se debe a que la incorporación de la información mecanística en la evaluación de riesgos muestra que las guías de inferencia pueden someterse a contrastación empírica. Resulta de ello que la influencia de los valores no cognitivos en la determinación de las guías de inferencia no está justificada en todos los casos, sino sólo en aquellos en que la contrastación empírica no es posible o no es conveniente dada su dificultad técnica o los costos asociados.

REFERENCIAS

Clewell H., 2005, ‘Use of mode of action in risk assessment: past, present, and future’, *Regulatory Toxicology and Pharmacology* 42/1, 3–14.

[2] Douglas utiliza la distinción de Laudan (2004) entre valores epistémicos y valores cognitivos.

Cranor, C. 1993, *Regulating Toxic Substances*, Nueva York, Island Press.

Cranor, C. 1997, 'The Normative Nature of Risk Assessment: Features and Possibilities', *Risk: Health, Safety and Environment* 8, 123-136.

Cranor, C. 2004, 'Conocimiento experto y políticas públicas en las sociedades tecnológicas: En busca del apoyo científico apropiado para la protección de la salud pública', in Luján, J. L.; Echeverría, J. (eds.), *Gobernar los riesgos: Ciencia y valores en la sociedad del riesgo*, Madrid, Biblioteca Nueva, 99-141.

Cranor, C. 2011, *Legally Poisoned: How the Law Puts Us at Risk from Toxicants*, Cambridge, Mass., Harvard University Press.

Douglas, H. 2006, 'Norms for values in scientific belief acceptance', *Contributed paper-20th PSA 2006 Vancouver*, 2-14.
[http://philsci-archive.pitt.edu/3024/1/Values_Beliefs_PSA_06_Final.pdf]

Douglas, H. 2009, *Science, Policy, and the Value-Free Ideal*, Pittsburgh, University of Pittsburgh Press.

Laudan, L. 2004, 'The Epistemic, the Cognitive and the Social', in Machamer, P. K.; Wolters, G. (eds.), *Science, Values and Objectivity*, Pittsburgh, University of Pittsburgh Press, 14-23.

López Cerezo, J. A., Luján, J. L. 2000, *Ciencia y política del riesgo*, Madrid, Alianza.

Luján, J. L. 2005, 'Metascientific analysis and methodological learning in regulatory science', in González, W. (ed.), *Science, Technology and Society: A Philosophical Perspective*, A Coruña, Netbiblo, 83-106.

Shrader-Frechette, K. 1994, *Ethics of Scientific Research*, Lanham, Rowman & Littlefield.

Shrader-Frechette, K. 2001, 'Radiobiological hormesis, methodological value judgments, and metascience', *Perspectives on Science* 8, 367-379.

Shrader-Frechette, K. 2004, 'Using metascience to improve Dose-Response Curves in biology', *Philosophy of Science* 71, 1026-1037.

United States Environmental Protection Agency 2003, *Guidelines for Carcinogen Risk Assessment*, Washington, DC, Risk Assessment Forum, U.S. Environmental Protection Agency. EPA/630/P-03/001A.

Wandall, B. 2004, 'Values in science and risk assessment', *Toxicology Letters* 152, 265-272.

White, R. H.; Cote, I.; Zeise, L.; Fox, M.; Dominici, F.; Burke, T. A.; Paul D. White, P. D.; Hattis, D. B.; Samet, J. M. 2008, 'State-of-the-Science Workshop Report: Issues and Approaches in Low-Dose—Response Extrapolation for Environmental Health Risk Assessment', EHP: *Environmental Health Perspectives*.
[ehp03.niehs.nih.gov/article/info:doi/10.1289/ehp.11502]

EL TECNOFEMINISMO DE JUDY WAJCMAN. ALCANCE TEÓRICO Y POLÍTICO

AMPARO ROMERO
Universidad de La Laguna

ABSTRACT: As a result of the development in information and communication technologies, the world is experiencing a great transformation, which is referred by some as a new type of society: the information society or the knowledge society. IT is certainly making our lives easier, but more than the strictly technical approach, its relevant aspect lies in its capacity to modify social relationships. Accordingly, Internet has not only become the paradigm of new information and communication technologies, but also the model of social relationships in the context of post modernity. Feminists have not remained at a standstill in the event of this new technological revolution, and new feminist movements that reflect upon the relationships between technological development and genderized social structures have made an appearance as a result of surfing the net. Movements such as cyberfeminism and technofeminism try to give account of how technological development has affected the construction of gender identity, taking special notice on the Internet, being this space considered the paradigm of New Technologies of Information and Communication.
KEYWORDS: Gender, Technology, Cyberfeminism, Technofeminism.

Ahora que la carrera espacial ha quedado obsoleta, hoy que la aventura del espacio ha sido sustituida por los viajes virtuales, los movimientos feministas se plantean un nuevo reto: la conquista del ciberespacio. El feminismo ha sabido adaptarse a lo largo de la historia a los diferentes contextos. Como ejemplo de este hecho, tenemos el gran y variado número de corrientes feministas que podemos enumerar en la actualidad. Desde las corrientes dominantes en las décadas de los setenta y ochenta, como el feminismo cultural, el feminismo liberal o el feminismo socialista, hasta las corrientes más recientes, como el ciberfeminismo o el tecnofeminismo, la lucha de las mujeres no se ha caracterizado en ningún momento por la homogeneidad de pensamiento.

Las feministas no han permanecido impasibles ante la nueva revolución tecnológica, y el nuevo tipo de relaciones que las Nuevas Tecnologías de la Información y la Comunicación han generado sirven de inspiración a autoras ciberfeministas como Sadie Plant para hacer resurgir imágenes tan sugerentes como la red, paradigmática del tipo de relación social actual que podemos hallar de manera especial en internet. Esta nueva conquista es, además, mucho mejor que la del espacio puesto que todos y todas podemos acceder a ella. Todas las personas podemos disfrutar de las ventajas de internet: trascender el tiempo y el espacio y escapar de las limitaciones del cuerpo físico, escapar de la “carne” como predica la ciberfeminista Sadie Plant, en un medio supuestamente democrático y no jerarquizado.

La entrada de las teóricas feministas a este tipo de análisis sobre el desarrollo tecnológico no ha resultado tarea fácil, fundamentalmente por un motivo: los estudios sociales de la ciencia y la tecnología se han mostrado reticentes a la aplicación de los análisis de la teoría feminista, protagonizados por el concepto de género, sobre sus objetos de estudio. Esto ha dado como resultado lo que Judy Wajcman denomina una *ceguera de género*[1] en los estudios de la tecnología de la corriente dominante, tanto en los más recientes como por ejemplo la teoría del actor-red, como en las corrientes más tempranas.

La actitud feminista hacia el desarrollo tecnológico no se ha caracterizado por la uniformidad de pensamiento, pero podemos realizar una división entre las posturas pesimistas/distópicas y las posturas optimistas/utópicas.

En el feminismo de las décadas de los 70 y 80 la tendencia general de las feministas con respecto a la tecnología fue la del pesimismo, tendencia que condujo directamente a la distopía. Según la visión distópica de la tecnología, ésta tiene un carácter patriarcal y su objetivo es dominar y explotar a las mujeres.

Actualmente, se percibe un clima generalizado de optimismo hacia el desarrollo tecnológico dentro y fuera de los movimientos feministas, ejemplificado en Internet como paradigmático del nuevo tipo de sociedad. Las nuevas tecnologías de la información y la comunicación suponen para muchas personas una ruptura con el carácter de los desarrollos tecnológicos anteriores. En palabras de Judy Wajcman:

> "Para muchas personas, la sociedad global de la información, que se caracteriza por la comprensión del espacio y del tiempo, marca toda una nueva época en la condición humana"[2]

Fuera de los análisis feministas, el optimismo hacia los nuevos desarrollos tecnológicos se encarna en autores como Manuel Castells, que opinan que Internet no está produciendo sólo una revolución tecnológica, sino también una revolución social. En su obra "La Sociedad Red" (2002) , Castells afirma que internet es la base para la creación de nuevas formas de relaciones sociales que nos ayudarán a recuperar los valores comunitarios. Internet es, en definitiva, la solución para el individualismo y la desintegración social.

No hay duda acerca de las ventajas de las nuevas tecnologías de la información y la comunicación en un mundo globalizado que necesita establecer redes de comunicación globales. Ahora bien, estas ventajas no pueden hacernos olvidar que las nuevas tecnologías, igual que las

1 Wajcman 2004, 66

2 Wajcman 2004, 10

anteriores, son creadas en un contexto social que las moldea atendiendo a las relaciones sociales que tienen lugar en dicho contexto, relaciones sociales que están mediadas, entre otras cosas, por el género. No se debe perder de vista el hecho de que la ciencia y la tecnología también son fuente de peligros y de nuevas desigualdades. En este sentido, las TIC's han acarreado, junto con un buen número de beneficios una nueva fuente de desigualdad que tiene lugar cuando las personas quedan al margen del desarrollo tecnocientífico y, por lo tanto, quedan excluidas de los ámbitos de decisión y de poder. Las reflexiones acerca de este tema se sitúan en lo que se denomina la "brecha digital", concepto que hace referencia a los colectivos de la población incluidos o excluidos de la sociedad de la información[3], es decir, al acceso desigual de las personas a las TIC's. Desde los estudios feministas se está comenzando a hablar de una "segunda brecha digital" que ha ampliado sus análisis del acceso a los usos. Esta brecha afecta de manera especial a las mujeres, puesto que, a pesar de que el número de usuarias de Internet ha crecido en los últimos años igualándose, en muchos casos, al de los hombres, los usos que realizan las mujeres son más restringidos y requieren menos destrezas tecnológicas[4].

A pesar de estos análisis y como ya se ha comentado, actualmente hallamos en el seno del feminismo un optimismo hacia las nuevas tecnologías, optimismo que conduce en la mayoría de los casos a la utopía.

Fue aproximadamente en la década de los 90 cuando comenzaron a surgir las visiones realmente optimistas en el seno del feminismo acerca de las consecuencias del desarrollo tecnológico en las vidas de las mujeres. Este punto de inflexión se debe a la obra de Donna Haraway, protagonizada por la figura del cyborg[5]. Desde el cyborg de Donna Haraway hasta los códigos binarios y las redes de la ciberfeminista Sadie Plant han transcurrido casi 30 años, suficientes para elevar el ciberespacio a la categoría de terreno de lucha idóneo para las mujeres del siglo XXI.

El ciberfeminismo comenzó en Adelaide (Australia) como un movimiento artístico de mujeres cuyo objetivo consistió en subvertir la identidad femenina en la red, espacio idóneo para esta tarea por su carácter flexible y abierto. Dicho grupo se denominó VNS Matrix y escribió el primer manifiesto ciberfeminista titulado "Manifiesto ciberfeminista para el siglo XXI" en 1991 que constituyó un homenaje al cyborg de Haraway. El término "ciberfeminismo" fue acuñado por Sadie Plant en 1995, dos años antes de la publicación de su obra "Ceros + Unos. Mujeres digitales + la nueva

3 Castaño 2008, 15

4 Castaño 2008, 29

5 Haraway 1985

tecnocultura" (1997), obra que puede ser considerada actualmente el soporte del fundamento teórico del movimiento ciberfeminista.

El ciberfeminismo es una corriente heterogénea y esto dificulta su definición. Las mismas ciberfeministas están en contra de los intentos por definir el movimiento y se decantan por decir lo que no es antes que lo que es. Durante el primer encuentro Ciberfeminista Internacional, que tuvo lugar en Alemania en septiembre de 1997, se escribieron las "100 antítesis" de lo que no es el ciberfeminismo. Algunas de ellas son

> "...el ciberfeminismo no es una fragancia, el ciberfeminismo no es una ideología, el ciberfeminismo no es aburrido, el ciberfeminismo no es una institución, el ciberfeminismo no es una teoría, el ciberfeminismo no es una práctica, el ciberfeminismo no es esencialista, el ciberfeminismo no es natural"[6]

Cien negaciones para realizar una "no-definición" del movimiento. Cien negaciones que han logrado que el ciberfeminismo carezca de una definicición certera y de un marco teórico consistente.

Teniendo en cuenta lo expuesto anteriormente, la descripción de ciberfeminismo que se realiza en el presente trabajo gira en torno al pensamiento de Sadie Plant como la representante teórica principal del movimiento, concretamente en torno a su obra "Ceros + Unos. Mujeres digitales + la nueva tecnocultura" (1997), en la que la filósofa reconstruye la identidad femenina asociándola con la informática, de manera concreta, y con las tecnologías que trabajan en red, de manera general. Esta nueva identidad se construye a partir de la asociación de la informática con la imagen de la red y la actividad de tejer, tradicionalmente feminizada. Para esto, Sadie Plant recurre a Ada Lovelace, considerada hoy la primera programadora informática, que recurrió a la idea de las tarjetas perforadas del telar de Jacquard. Con esto cierra Plant el círculo: identidad femenina, ordenadores, actividad de tejer, comunicación en red.

Las nuevas tecnologías, pues, son esencialmente liberadoras para las mujeres, puesto que su apertura y flexibilidad permiten la libre creación de la identidad y, además, las nuevas tecnologías de la información y la comunicación, afirma Sadie Plant, tienen un carácter femenino puesto que su uso requiere de las habilidades que, tradicionalmente, han sido típicas de las mujeres. La capacidad de tejer, de unir hilos y redes, es extrapolada por el ciberfeminismo a las comunicaciones y formas de relacionarse que se establecen en las nuevas tecnologías de la información, especialmente en el ciberespacio.

6 100 anti-theses. Cyberfeminism is not... en http://www.obn.org/cfundef/100antitheses.html

Las nuevas tecnologías de la información, afirma Sadie Plant, no sólo implican la subversión de la identidad masculina, sino también la creación de múltiples identidades innovadoras que pueden acabar con los roles de género tradicionales. Escapar de la "carne", escapar del determinismo biológico que encierra el género en un cuerpo físico se convierte en el punto alrededor del cual gira el pensamiento utópico ciberfeminista, que reacciona frente a los movimientos feministas anteriores, consideradas por el ciberfeminismo como pesimistas, tecnófobos y esencialistas. Pesimistas porque, desde su punto de vista, mantienen una postura distópica con respecto al tema de la relación entre las mujeres y las máquinas; tecnófobos porque el pesimismo las lleva a alejar a las mujeres de los artefactos tecnológicos; esencialistas porque encierran el género y la identidad de los individuos en los cuerpos biológicos. La misma Sadie Plant afirma que este tipo de feminismo, que ella denomina feminismo clásico, ya ha desaparecido y su lugar debe ser ocupado por el ciberfeminismo.

El optimismo ciberfeminista con respecto a las nuevas tecnologías de la información y la comunicación acarrea una serie de consecuencias, tanto en el ámbito teórico como en el político, que dificultan los análisis críticos con perspectiva de género del impacto social de dichas tecnologías y, por tanto, también obstaculizan las acciones políticas encaminadas a la igualdad de género.

En el ámbito teórico, el ciberfeminismo adolece en general de una visión crítica con perspectiva de género que sea capaz de dilucidar hasta qué punto un producto surgido en una determinada sociedad con unos valores concretos –en este caso patriarcales- puede ser el instrumento idóneo para la emancipación de las mujeres o si, por el contrario, la misma tecnología, como producto sociotécnico[7], constituye un reflejo de las relaciones sociales mediadas por el género. En el mismo ciberespacio conviven visiones contradictorias acerca de las mujeres y de los hombres. Navegando por internet, hallamos un buen número de páginas web feministas, pero también podemos encontrar páginas que refuerzan claramente los estereotipos de género tradicionales como, por ejemplo, las páginas de belleza y salud dirigidas a las mujeres. Con este ejemplo se percibe de manera clara el hecho de que los valores y estereotipos en los que se fundamenta las relaciones asimétricas entre los sexos se trasladan, frecuentemente, al ciberespacio y éste pasa a ser, gracias a su enorme poder de difusión, un instrumento que perpetúa y aumenta las desigualdades nombradas.

En el ámbito político vemos como, a pesar de su afán por alejarse del esencialismo que caracterizó a los feminismos de las décadas de los 70 y 80, Sadie Plant sigue ubicada en él tanto en

7 Wajcman 2004

lo referente a la feminidad, como en la categoría de las nuevas tecnologías de la información como esencialmente liberadoras para las mujeres, aspecto que la conduce directamente al determinismo tecnológico. El esencialismo referido a las nuevas tecnologías como liberadoras para las mujeres ve innecesario cualquier tipo de activismo político puesto que, si aceptamos la premisa de que las nuevas tecnologías, por sí mismas, acabarán con la condición negativa de las mujeres, ¿con qué objetivo se llevarán a cabo las acciones políticas? En efecto, como afirma acertadamente Judy Wajcman, si pensamos que la tecnología es inherentemente femenina no es necesaria ninguna acción política.[8]

No cabe duda de que las acciones políticas son importantes, pues fuera de la comunidad virtual libre y abierta que predican las ciberfeministas, existe un mundo real constituido por comunidades reales[9] en las que las diferencias de poder están bien claras y nos llegan mediadas por diversas categorías, entre ellas el género.

En este sentido, es necesario un marco teórico que permita realizar un análisis completo de las nuevas tecnologías de la información y la comunicación y que posibilite el salto del ámbito teórico al activismo político sin dificultades. Esto lo podemos hallar en la alianza que se ha establecido entre constructivismo y feminismo.

La alianza entre constructivismo y feminismo surge por la aportación realizada desde el segundo movimiento al primero del género como una categoría útil para analizar las relaciones recíprocas entre tecnología y sociedad. Judy Wajcman ha otorgado el nombre de "tecnofeminismo" a dicha alianza y lo define de la siguiente manera:

> "Cierto tecnofeminismo emergente concibe una relación mutuamente conformadora entre género y tecnología, en la que la tecnología es, al mismo tiempo fuente y consecuencia de las relaciones de género"[10]

Las nuevas tecnologías de la información y la comunicación deben ser analizadas desde un punto de vista materialista y constructivista en el que aparezca la perspectiva de género y en el que se tenga en cuenta que la asociación de la tecnología con la masculinidad es una *construcción social*[11].

La tecnología es, para Judy Wajcman un producto sociotécnico. Este término hace referencia a la concepción de la tecnología como resultado tanto de procesos tecnológicos como de procesos sociales. La tecnología no es neutral, sino que su creación y desarrollo responde a las estructuras

8 Wajcman 2004, 114

9 Sanz González 2006

10 Wajcman 2004, 161

11 Alemany 1999, 82

sociales en las cuales surge. De esta manera, observamos que existe también un proceso de intercambio entre el género y la tecnología en el que

> "la tecnología no sólo afecta a las relaciones y a la definición de género, sino que en toda innovación tecnológica se produce una renegociación de las relaciones y una articulación de las identidades de género que van a ser performadas con el uso de ese artefacto"[12]

Los símbolos y los significados que aparecen en los artefactos tecnológicos no existen de manera previa, sino que se van configurando a lo largo de todo el proceso constructivo del producto. Esta postura permite el alejamiento del pesimismo, al admitir el carácter cambiante y heterógeneo del desarrollo tecnológico, y también del peligro de esencializar el género, al contemplar la posibilidad de que éste se transforme conjuntamente con la tecnología. Al mismo tiempo, nos otorga la posibilidad, aplicando la perspectiva de género, de explicitar hasta qué punto los productos tecnológicos están influenciados por los significados y las relaciones de género.

En definitiva, preguntarnos por las ventajas que las nuevas tecnologías ofrecen a las mujeres es muy importante, si tenemos en cuenta que la tecnología domina todos los ámbitos de nuestra vida, pero también debe realizarse el esfuerzo de alejarse del marco normativo y de lo que debería ser y realizar análisis constructivistas y materialistas que conciban a la tecnología como un producto sociotécnico. En el caso que nos ocupa, aplicar la perspectiva de género es fundamental, pues es la que nos ayuda a dilucidar los estereotipos y roles tradicionales de género que, aún hoy, podemos hallar en los desarrollos y aplicaciones de la tecnología. En este sentido y partiendo de la premisa de que el contexto en el que se desarrolla la tecnología influye en la misma, debemos concebir que la tecnología es flexible y, por lo tanto, que puede tener tanto efectos beneficiosos como efectos perniciosos. En palabras de Judy Wajcman:

> "La propia tecnología es plástica y, por consiguiente, una misma tecnología puede tener efectos contradictorios, dado que el contexto y las relaciones sociales de su utilización inciden en la misma"[13]

REFERENCIAS

Libros

Wajcman, J. 2004, *Tecnofeminismo*, Madrid, Ed. Cátedra

Castaño, C. 2008, *La segunda brecha digital*, Madrid, Ed. Cátedra

Plant, S. 1998, *Ceros + Unos. Mujeres digitales + la nueva tecnocultura*, Barcelona, Ed. Destino

12 Wajcman 2004, 162

13 Wajcman 2004, 112

Castells, M. 2002, *La Era de la Información. Vol. I: La Sociedad Red*, México, Distrito Federal, Siglo XXI Editores

Artículos en libros

Carme Alemany, M. 1999, *Tecnología y género: la reinterpretación de la tecnología desde la teoría feminista* en M.J. Barral, C. Magallón, C. Miqueo, M.D. Sánchez (eds.): *Interacciónes ciencia y género*, Barcelona, Ed. Icaria

Artículos en revistas digitales

Sanz González, V. 2006, *Las tecnologías de la información desde el punto de vista del género: posturas y propuestas desde el feminismo* en www.isegoria.com

LAS PATENTES, NUEVAS FUENTES PARA LA HISTORIA DE LA CIENCIA Y LOS ESTUDIOS CTS

ANA ROMERO DE PABLOS
Instituto de Filosofía, CCHS-CSIC

Abstract: The aim of this paper is to construct history of science based on patents archival for examining the interaction between science, technology and society. I argue that by this practices patents became scientific objects and subjects that contribute to epistemology.
Key words:
Patents, history of science, circulation of knowledge, regulation, STS.

El objeto de este trabajo es mostrar las patentes como fuentes e instrumentos claves para la historia de la ciencia y para los estudios que analizan la interacción ciencia tecnología y sociedad. Asimismo se analiza su relevancia como objetos y sujetos científicos que influyen en el análisis histórico-epistemológico de la ciencia. Hoy día son múltiples las vías por las que viaja la información y el conocimiento científico y tecnológico. Congresos científicos, publicaciones, tesis doctorales o exposiciones son medios de acceso al conocimiento por todos conocidos. Por el contrario, las patentes quedan fuera de este espacio, algo que no deja de resultar curioso pues son documentos que por ley están obligados a ser públicos.

Las ciencias, como todo conocimiento, son productos sociales y por ello no pueden ser explicadas o analizadas con una única voz, la de los científicos y los expertos. En su desarrollo intervienen múltiples factores políticos, sociales y económicos, por lo que, además del conocimiento, tenemos que tener en cuenta las prácticas. Las nuevas formas de obtención de conocimiento implican nuevos actores, nuevas demandas, nuevos procedimientos, nuevas estructuras, nuevas formas de evaluarlo y de gestionarlo y también, nuevas formas de participación social. Las patentes ayudan a entender la ciencia como producto social e intelectual, a comprender la dinámica social, quién la produce, con qué fines, cuáles son sus prácticas, con qué costes, cuáles son los beneficiarios, etc. Por último, el estudio con patentes alimenta la conciencia social sobre ciencia y tecnología. Una base social motivada y crítica amplía la percepción favorable a la investigación y la innovación, y propicia la modernización y el cambio.

La historiografía muestra que son muchos los trabajos que utilizan las patentes como objeto de estudio. Gran parte de los estudios históricos sobre patentes en los que se ha reflexionado sobre los sistemas de I+D se han centrado en la repercusión que han tenido las normativas estatales, y también en cómo la generación de nuevo conocimiento da lugar a su vez a instituciones propias para administrarlo (Mowery, et al. 2001; Mowery y Ziedonis 2002; Metlay 2006). En otros, la patente se convierte en una herramienta para conocer cómo se gesta y construye el conocimiento científico y tecnológico (Romero de Pablos 2005 y 2011; Romero de Pablos y Azagra-Caro 2009), los nuevos espacios de interacción que se generan con la introducción de nuevos actores (Slinn 2008), las necesarias negociaciones para proteger y delimitar espacios para que al tiempo puedan coexistir (Gaudillière 2008), y las nuevas prácticas y dinámicas que introducen las modificaciones legales y la repercusión en la industria (Guagnini and Inkster 2004; Cassier 2008). Estos trabajos indican que en el origen de las patentes hay también otros motivos que en ocasiones pueden incluso pesar más que la protección del conocimiento. Hay patentes con las que lo que se persigue es divulgar y publicitar una invención, buscando más una legitimación académica, empresarial, industrial y social que la protección de la invención. Otras que vienen motivadas por estrategias diseñadas por las propias empresas o instituciones donde se generan. Algunos sectores utilizan las patentes para llamar la atención sobre desarrollos determinados y localizar y dirigir de esta manera las inversiones, acotando espacios. Estos pocos ejemplos muestran que el panorama es amplio por lo que parece muy reduccionista ver las patentes sólo como indicadores de la innovación. Si a esto añadimos, como nos muestra la historia, que de forma constante hay que definir la naturaleza de nuevos objetos patentables, repensar los marcos legales, etc., (Guagnini and Inkster 2004), parece razonable sumar a los modelos económicos utilizados por los economistas de la innovación, otras variables importantes.

Junto al papel que juegan las patentes en el proceso de estandarización, normalización y control del conocimiento, son también instrumentos o herramientas útiles y necesarias para conocer cómo éste se genera y circula. El estudio de las patentes puede dar nuevas visiones sobre cómo viaja el conocimiento y cómo se gestiona, además de aportar novedades sobre la extensión industrial y tecnológica de las prácticas científicas o sobre la comunicación entre científicos y público. Por ello es necesario abordar el análisis de las patentes teniendo presente los contextos en los que surgen y dónde se desarrollan (Romero de Pablos 2011).

Las patentes son portadoras de nuevo conocimiento y también son capaces de generarlo. La incorporación de las patentes a los estudios CTS favorece expresiones micro de lo macro cambiando nuestra forma de ver el mundo y generando nuevos espacios. Las patentes y las historias que contienen afectan de forma comprometida no solo a los científicos implicados, sino que también a una sociedad como la actual donde todos los aspectos de la cultura moderna están impregnados de ciencia y tecnología. Los estudios con patentes revelan la existencia de múltiples

procesos donde están presentes la política, la economía, la ciencia, la tecnología y también los temas de género.

Obviando el uso de patentes con fines legales, en principio se pueden establecer dos ámbitos claros de aplicación de las mismas: el primero donde hay un uso inmediato de la tecnología patentada como fuente de ayuda imprescindible al I+D, y el segundo, donde un uso más reposado, analítico y a más largo plazo, permite que la industria y las administraciones públicas planifiquen actividades futuras y opten por estrategias a seguir.

La información contenida en las patentes permite estudiar el estado de la técnica de un sector tecnológico determinado, seguir la evolución tecnológica que se ha producido -cuáles han sido las tendencias, cómo han sido los tiempos-, y cómo se ha producido la transferencia de tecnología -cómo ha viajado, cómo se han establecido las relaciones entre las partes interesadas, cómo se ha producido la aportación del capital, si esta transferencia de tecnología ha implicado formación de personal, si se han producido modificaciones en la forma de hacer ciencia, en los métodos de gestión, las instituciones u organismos implicados.

Las patentes trascienden los espacios (disciplinares y geográficos) donde tienen lugar los procesos administrativos, y también los espacios propios de actuación de los científicos y tecnólogos.

Las patentes abren así nuevas líneas de trabajo para el análisis histórico-epistemológico en historia de la ciencia y para los estudios que analizan la interacción ciencia tecnología y sociedad.

REFERENCIAS

Cassier, M. 2008, "Patents and public health in France. Pharmaceutical patent law in-the-making at the patent office between the two world wars". *History and Technology* 24 (2), 135-151.

Etzkowitz, H. 1994, "Knowledge as property: The Massachusetts Institute of Technology and the debate over academic patent policy", *Minerva* 32 (4), 383-421.

Gaudillière, J.P. 2008, "Professional or industrial order? Patents, biological drugs, and pharmaceutical capitalism in early twentieth century Germany" *History and Technology*; 24 (2), 107-133.

Guagnini, A. and Inkster, I. 2004, "Patents in History" *History of Technology*. Volume Twenty-Four.

Mowery, D. C. et al. 2001, "The growth of patenting and licensing by U.S. universities: an assessment of the effects of the Bayh-Dole Act of 1980", *Research Policy* 30, 99-119.

Mowery, D. C.; Ziedonis, A. 2002, "A. Academic patent quality before and after the Bayh-Dole Act in the United States", *Research Policy* 31, 399-418.

Metlay, G. 2006, "Reconsidering renormalization: stability and change in 20th-century views on university patents", *Social Studies of Science* 36 (4), 565-597.

Romero de Pablos, A. 2005, "Gobernanza y gestión del conocimiento: las patentes, un instrumento de estudio", *Arbor* 715, 333-350.

Romero de Pablos, A. 2011, "Regulation and circulation of knowledge: Penicillin patents in Spain", *Dynamis* 31 (2), 363-383.

Romero de Pablos, A. y Azagra-Caro, J. M. 2009, "Internationalization of patents by Public Research Organizations: from a historical and economic perspective". *Scientometrics* 79 (2), 329-340.

Slinn, J. 2008, "Patents and the UK pharmaceutical industry between 1945 and the 1970s", *History and Technology* 24 (2), 191-205.

LOS IMAGINARIOS COLECTIVOS EN LA REPRESENTACIÓN SOCIAL DE LA CIENCIA

MIGUEL ZAPATA
UNAM

> La nave de nuestra imaginación, perfecta como un copo de nieve y sutil como un diente de león, nos llevará a mundos de ensueños y grandes verdades.
>
> Carl Sagan. *Cosmos*

ABSTRACT: The scientific culture in a society can be measured by the understanding and assessment that the whole citizenship has about the practices, knowledge and scientific productions, such as the influence that science has over the daily decisions of the population. The studies about social representation are an important theoretical framework to map the scientific culture. This works aims to establish a classification about the social representations based on the rational aspects as well as the imaginary ones that condition the beliefs and attitudes about the science. With this proposal will become apparent the importance having the generation of metaphors, symbols and the other imaginary elements in the configuration of the social representation. Also can be include, against the perspectives that considerate that these symbolic devices just simplify and deform the scientific contents that are ineludible elements in order to produce legitimist knowledge and to develop a healthy scientific culture.
KEYWORDS: Ideology, metaphor, scientific culture, social imaginary, social representation

Los estudios sobre representación social han constituido uno de los marcos teóricos más eficaces y prolíficos para evaluar la cultura científica de una sociedad. Sin embargo, existen pocos trabajos que hayan abordado un análisis sobre los distintos tipos de representaciones sociales de la ciencia. El objetivo de la presente ponencia es establecer una clasificación que colabore en la subsanación de dicha carencia. Para llevar a cabo esta tarea se partirá del análisis de una propuesta taxonómica elaborada por Olivé (2007, 143-167), basada en la racionalidad y los intereses que subyacen a las creencias que distintos grupos comparten respecto a la ciencia. Este criterio permite diferenciar, al menos, tres grandes clases de representaciones sociales:

a) Objetivas. Son creencias sobre la ciencia que encuentran justificación en las mejores razones disponibles. Ejemplos de estas representaciones pueden ser: "La ciencia requiere de un sistema educativo para su continuidad" o "la ciencia contemporánea es una actividad colectiva". La

objetividad de estas afirmaciones queda avalada por cualquier agente racional que sea capaz de plegar sus creencias ante la fuerza de los argumentos que den mejor cuenta de los hechos referidos.

b) Constitutivas de la ciencia. Son las opiniones y actitudes mantenidas dentro de la comunidad científica que, si bien son susceptibles de controversia por carecer de razones concluyentes, determinan el sentido y la autocomprensión de sus propias prácticas. Un ejemplo lo constituye la extendida opinión entre los científicos de que el objetivo principal de su actividad es la búsqueda de la verdad. Al condicionar la actividad de los científicos que las sostienen, estas representaciones adquieren carácter performativo.

c) Ideológicas. Son representaciones que no tienen correspondencia con los hechos y que carecen de justificación racional. Ejemplos de este tipo son: "Todos los sistemas tecnocientíficos tienen objetivos filantrópicos" o "hay una incompatibilidad absoluta entre la mentalidad científica y el espíritu religioso". Como se podrá ver más adelante, con esta clase de afirmaciones se pretende conseguir aceptación pública para legitimar ciertas prácticas científicas.

Esta clasificación aporta dos grandes ventajas. Por un lado, contribuye a la evaluación de las representaciones sociales según su sometimiento a los mejores argumentos disponibles. Por otro, hace explícita la existencia de un sustrato ideológico en algunas de las creencias que mantienen tanto la comunidad científica como la opinión pública, desvinculándose así de interpretaciones que tradicionalmente han considerado la ciencia y la ideología como dos esferas antagónicas; falsa una, correcta la otra (Althusser, 1970).

A pesar de estas ventajas, la existencia de representaciones sociales que sin encontrar apoyo racional resultan ajenas a intereses ideológicos obliga a una revisión de la tipología presentada. Si se considera una de las características centrales del concepto de ideología, las representaciones ideológicas de la ciencia deberían entenderse como opiniones infundadas cuyo objetivo "es la promoción o el mantenimiento en el poder de algún grupo" (Villoro, 1985). Por tanto, no podrían subsumirse dentro de esta clase las representaciones no objetivas y desinteresadas. Sin embargo, el mayor inconveniente de esta clasificación es la preeminencia otorgada a los aspectos racionales en detrimento de los simbólicos. Esta desconsideración, lejos de resultar irrelevante, puede tener implicaciones negativas para las evaluaciones de la cultura científica y la eficacia de los proyectos destinados a su mejora. Por eso un análisis de las representaciones sociales debería integrar lo simbólico, aunque para ello ha de establecerse con precisión el sentido que se le dará a dicho término. Para los fines del presente trabajo, los estudios sobre el imaginario colectivo ofrecen el mejor marco teórico para definirlo.

La noción de imaginario colectivo hace referencia a la facultad que produce la urdimbre simbólica por la que se constituye y organiza el mundo social. Dentro de esta línea interpretativa

existen grandes diferencias teóricas. Algunos, siguiendo las tesis de Durand (2006), lo entienden como una proyección arquetípica proveniente del inconsciente; mientras que otros, como Castoriadis (1988), defienden la existencia de una capacidad radical que se nutre de un magma simbólico inagotable y no condicionado por estructura antropológica alguna. Sean cuales sean las fuentes del imaginario, lo que aquí interesa recoger es la idea de un dispositivo simbólico que condiciona las formas de habitar y comprender el mundo al generar "esquemas colectivos de interpretación" que los individuos interiorizan y por los que moldean su conducta (Bazcko, 1999, 30). Tomando esto en consideración, puede definirse el imaginario colectivo de la ciencia como:

1) el dispositivo que genera el arsenal simbólico (metáforas, analogías, ficciones…) a partir del cual se constituyen las creencias, imágenes, ideas o actitudes respecto a la ciencia 2) interiorizadas y compartidas de manera colectiva por los miembros de algún grupo, 3) que abren determinados senderos de interpretación o impulsan prácticas colectivas.

Con esta noción y tomando como referencia la primera clasificación que se presentó, a continuación se propone la siguiente tipología de representaciones sociales sobre la ciencia:

a) Representaciones Ideológicas

Si bien la especificidad de este tipo de representaciones reside en el interés y la falta de objetividad, no puede ignorarse la importancia que en ellas adquiere la utilización de elementos imaginarios. Cuando biotecnólogos manifiestan que sus acciones se asemejan a una operación quirúrgica, o cuando desde la televisión se proyecta una bella imagen de la Tierra tomada desde el espacio para publicitar la conciencia medioambiental de una empresa petroquímica; se está poniendo en marcha una batalla simbólica cuyo objetivo es configurar representaciones sociales con actitudes favorables hacia ciertas prácticas. Si a esto se le añade el poder de las metáforas y las imágenes para influir en las acciones colectivas, se percibe claramente la gravedad que supone el bombardeo simbólico con fines ideológicos. Al respecto, un reciente estudio de psicología social ayudará a comprender el gran poder que poseen las metáforas (Thibodeau, P, Boroditsky, L. 2011). En dicha investigación se expuso a los participantes la situación hipotética de una ciudad que sufre un elevado índice de criminalidad. Mientras a un grupo se le describió la delincuencia como si fuera un virus, a otro, en cambio, se le presentó como un monstruo. A continuación, cuando se les preguntó a los participantes por las medidas que propondrían para reducir el número de crímenes de la ciudad, se pudo apreciar el poder de la metáfora para condicionar las formas de interpretar y abordar la situación. Los que recibieron la metáfora del virus se mostraron más proclives a proponer medidas encaminadas a investigar la raíz del problema o a desarrollar programas educativos. En cambio, el grupo al que se había presentado la metáfora del monstruo manifestó preferencias por

otro tipo de medidas, como la contratación de un mayor número de policías o la construcción de más cárceles.

La conclusión que se desprende de esta investigación es que las metáforas tienen una gran capacidad para influir en la toma de decisiones. Por este motivo, cuando nos enfrentamos a problemas colectivos que se interpretan a la luz del conocimiento científico disponible, como es el caso del cambio climático, se torna necesaria la realización de análisis que permitan detectar las imágenes ideológicas con las que se pretenden dirigir las decisiones. Esta labor de rastreo y crítica debería realizarse, especialmente, en aquellos medios que tienen mayor mayor influencia para la conformación de la opinión pública, como los libros de texto y los programas televisivos.

b) Representaciones constitutivas de la ciencia

Los dispositivos simbólicos de estas representaciones han determinado la autocomprensión de la empresa científica en momentos de especial relevancia para su desarrollo. Para aceptar esta idea no hay más que fijar la atención en los orígenes de la ciencia moderna y reconocer la importancia que tuvo la metáfora que transfiguró la naturaleza en máquina para la consolidación de los proyectos mecanicistas. Otro interesante ejemplo donde se manifiesta la importancia de lo simbólico puede encontrarse en la subordinación de las matemáticas a los imaginarios de las culturas en que ésta se ha desarrollado. Sobre esta cuestión, Lizcano (2009) ha defendido con argumentos convincentes la idea de que la incapacidad de los matemáticos griegos para concebir la negatividad y el número 0 se debió a los presupuestos metafísicos que apuntalaban su cultura. Según esta tesis, los principios de identidad, no contradicción y tercio excluso habrían dificultado el camino para la comprensión de los números negativos hasta Diofanto; mientras que la tradición eleática habría obstaculizado el descubrimiento de una entidad matemática apropiada para referirse a la ausencia. La cultura china, en cambio, con un imaginario impregnado por la dualidad del *yin* y el *yan*, pudo entender la negatividad como un ámbito complementario de los números positivos y el número 0 a partir de la noción del *tao*, como un gozne donde los contrarios se anulan.

c) Representaciones objetivas

A pesar de estar basadas en los mejores argumentos disponibles, estas representaciones tampoco son inmunes a la colonización imaginaria de sus discursos. En estos casos, los símbolos son recursos epistémicos necesarios para la mejor interpretación posible de la realidad. Tradicionalmente la consideración de lo imaginario como un aspecto relevante para la comprensión de la ciencia ha quedado restringida a la divulgación. En este ámbito, las metáforas o imágenes que ejercen una función pedagógica sin pervertir el contenido científico siempre han sido entendidas

como un valioso recurso para acercar la ciencia a la sociedad, aunque esto supone correr el riesgo de elegir imágenes que, al contrario, oscurezcan o distorsionen el sentido de aquello que se quiere comunicar. A este respecto, una metáfora que ha demostrado ser útil para facilitar la comprensión de algunos fenómenos atmosféricos es la de efecto invernadero. Por el contrario, llamar al Bosón de Higgs partícula divina (Lederman, 1996) quizá sea un exceso retórico que puede llegar a ser perjudicial .

De cualquier forma, las representaciones objetivas imaginarias que integran esta clase nada tienen que ver con funciones simplificadoras que permitan hacer accesible a la opinión pública un conocimiento abstruso. Al contrario, son representaciones conformadas por símbolos útiles para la construcción de las teorías y la interpretación científica del mundo. Lejos de lo que pudiera parecer a primera vista, son muchas las representaciones de este tipo que se encuentran en la ciencia. La teoría evolucionista es especialmente rica en ejemplos, pues Darwin, considerado un maestro de las metáforas, no dudó en utilizarlas para desarrollar sus ideas sobre la evolución. Si alguien aceptara que la selección natural es una ley que rige la evolución de las especies, estaría asumiendo el potencial explicativo de una metáfora, la de selección. En este caso el elemento metafórico no se utiliza para trasladar al profano un contenido científico, sino que el propio científico la necesita para explicar al interior de la teoría un aspecto importante del proceso evolutivo. Entre las muchas representaciones objetivas con carga imaginaria que pueblan la ciencia pueden destacarse las que se refieren al concepto de información usado por la genética o las ciencias cognitivas, o, sin abandonar a Darwin, las que se apropian de la imagen de un árbol para interpretar la historia evolutiva de las especies. El éxito de tales metáforas en la interpretación científica del mundo, así como su validación según los criterios establecidos por las disciplinas que las acogen, las convierte en conceptos epistémicamente relevantes y "nos justifica racionalmente para depositar cierta confianza en su capacidad representativa" (Marcos, 2010, 352-354).

d) Representaciones imaginarias

En todas las representaciones analizadas hasta ahora se han encontrado componentes simbólicos; sin embargo, en cada una de ellas había algún otro aspecto que determinaba su especificidad: en las ideológicas, el interés; en las constitutivas, la performatividad; en las objetivas, los buenos argumentos. Por el contrario, en las representaciones que ahora llamamos imaginarias, lo simbólico es el elemento predominante de su configuración.

Un ejemplo de representación social imaginaria podría condensarse en la afirmación: "La Ley General de la Gravitación Universal fue descubierta gracias a la caída de una manzana." Esta creencia está bastante generalizada; sin embargo, si se pidiera a los que aceptan la veracidad de la

anécdota que dieran los motivos por los que creen que esto sucedió así, probablemente pocos serían capaces de responder algo concreto. De cualquier forma esto no es demasiado importante, pues a menos que se sea un historiador de la ciencia especializado, parece poco relevante saber si la anécdota la inventó Newton para explicar a su sobrina su teoría o si las palabras que ésta última contó a Voltaire son rigurosamente ciertas (Andrade, 1943). La imagen de la manzana cayendo en el huerto de Newton se ha desprendido de los hechos para formar parte del imaginario cultural. Por eso, aun cuando alguien tuviera sólidas pruebas de que la famosa caída no inspiró jamás al científico, no sería razonable que se pretendiera acabar por eso con un símbolo que fomenta en las nuevas generaciones curiosidad por el personaje y los introduce en las primeras nociones de la mecánica clásica.

Razones parecidas pueden aducirse para defender la existencia del resto de símbolos que pueblan la imagen de la ciencia. La presencia de un variado imaginario científico es un buen indicador de una cultura científica sólida, pues la creación, proliferación, expansión y aceptación de símbolos referidos a la ciencia sólo puede darse en una sociedad que tenga especial interés en ella. Por eso, para hacer una evaluación de la cultura científica también resulta pertinente analizar el papel que juegan en la configuración de la representación social los medios que más imaginario científico generan.

La literatura de ciencia ficción, al haber creado narrativas con una gran cantidad de referencias a la ciencia, siempre ha tenido una gran influencia para aumentar el interés del público por ésta. Sin embargo, y sin desdeñar las ficciones literarias contemporáneas, si hay algún medio que tenga en la actualidad la capacidad para conformar representaciones sobre la ciencia es la televisión. En un rápido zapping se encuentran muchos ejemplos. *Los Simpsons* han tenido un hueco en la revista *Nature* (Hopkins, 2007) por aludir regularmente a la ciencia y haber incluido en sus capítulos a gente como Jay Gould o Hawking. En *Futurama*, guionistas que tienen doctorados en física resuelven disparatadas tramas con teoremas matemáticos que después son comentados en blogs especializados en ciencia. *The Big Bang Theory* parece haber sido uno de los factores que influyó en el incremento de matrículas en las facultades de física al proyectar una imagen divertida de los investigadores. *Star Trek* ha servido de inspiración a físicos como Alcubierre, quien desarrolló su mecánica para tratar de explicar cómo puede viajar tan rápido la nave Enterprise (Alcubierre, 1994). Los problemas clínicos a los que se enfrenta el doctor House sirven para mostrar algunas de las controversias analizadas por los comités de bioética. La lista completa de programas sería demasiado larga, pero basta con este reducido número de casos para mostrar que la proliferación de creaciones ficticias, al hacer de la ciencia algo atractivo, agradable y cotidiano, influye positivamente en la conformación de la representación social y en la promoción de la

propia empresa científica. La conclusión que se desprende de esto es que los dispositivos imaginarios no ideológicos o epistémicamente válidos mejoran la percepción pública de la ciencia. Por esta razón, cualquier proyecto que promueva este tipo de imaginarios estará colaborando con el desarrollo de una cultura científica saludable.

REFERENCIAS

Alcubierre, M. 1994, 'The Warp Drive: Hyper-fast Transluminic within General Relativity', Classical and Quantum Gravity, 11(5), 73-77

Althusser, L. 1970, 'Idéologie et appareils idéologiques d'État. (Notes pour une recherche)', *La Pensée* 151.

Andrade, C. 1943,'Newton and the Apple', *Nature* 151, 84.

Baczko, B. 1991, *Los imaginarios sociales. Memorias y esperanzas colectivas*. Buenos Aires, Nueva Visión.

Castoriadis, C. 1988, *Los dominios del hombre: las encrucijadas del laberinto*, Barcelona, Gedisa.

Durand, G, 2006, *Las estructuras antropológicas de lo imaginario*, México, FCE.

Hopkin, M. 2007, 'Science in comedy: Mmm... pi', *Nature* 448, 404-405.

Lederman, L., Teresi, D, *La partícula divina,* Crítica, Barcelona, 1996.

Lizcano, E. 2009, *Imaginario colectivo y creación matemática*, Barcelona, Gedisa.

Marcos, A. 2010, *Ciencia y acción*. México, Fondo Cultura Económica.

Olivé, L. 2007, *La ciencia y la tecnología en la sociedad del conocimiento*, México, Fondo Cultura Económica.

Thibodeau, P, Boroditsky, L. 2011, 'Metaphors We Think With: The Role of Metaphor in Reasoning', *PLoS ONE* 6, 2.

Townsen, M. 2011, 'Big Bang Theory fuels physics boom', *The Guardian*, 6-11.

Villoro, L. 1985, 'Del concepto de ideología'. En Luis Villoro, *El concepto de ideología y otros ensayos*, 15-40, México, FCE.

Symposia / Simposios

LOGIC, HISTORY AND PHILOSOPHY OF LOGIC

LÓGICA, HISTORIA Y FILOSOFÍA DE LA LÓGICA

PRACTICE-BASED APPROACHES TO PHILOSOPHY OF LOGIC

JOSÉ FERREIRÓS, MARÍA JOSÉ FRÁPOLLI
(coordinadores)

In recent times quite a lot of attention has been devoted to the philosophy of mathematical practice (or practices), as attested by several collective volumes and the founding of a new association, APMP. At the same time, there have been some attempts to move along parallel lines regarding the philosophy of logic, like e.g. a workshop entitled "From practice to results in logic and mathematics" that took place in Nancy, France, 2010 (which has led to a special issue of *Philosophia Scientiae,* Vol. 16, nº 1). We intend to discuss a number of questions that arise naturally in connection with this matter. What is novel about practice-based approaches to the philosophy of logic? Should they focus on the practice of logic itself as an independent discipline, or on non-logical practices as well? Are such approaches coincident with pragmatist takes on logic? Can they contribute significantly to clarifying what logic is, and what it is about? In particular, could they make a contribution to the long-standing debate on the logical constants?

Practice-based approaches may simply be taken to *complement* traditional philosophy of logic, for instance by asking new questions about the cognitive underpinnings of inferential practices, the role of formal languages and other forms of representation (e.g. diagrams) in them, etc. They can also be proposed as *alternatives* to more traditional, essentialistic perspectives on logic and its foundations. Finally, they offer a natural ambient space for asking again old questions about the history and the future of logic, like: What is common to the different activities that have received the name 'logic' at different times and places?

PROSPECTS FOR PRACTICE-BASED PHILOSOPHY OF LOGIC*

JOSÉ FERREIRÓS
Universidad de Sevilla. Spain
josef@us.es

ABSTRACT: We explore prospects for practice-based approaches to logical theory, in particular the link between classical and intuitionistic logic and the inferential structure of traditional practices of representation & argument in science and mathematics. After discussing some key notions about practice, we outline the connection between representation practices and classical logic, and then consider a spectrum of actual practices followed or proposed by (real) scientists. Intuitionistic logic helps to clarify the potential of practice-based approaches for understanding pluralism, and to hammer some key points about the general thesis.
KEYWORDS: pragmatism, practice-based philosophy, mathematical practice, view of logic as model, logical constants, classical logic, intuitionistic logic, pluralism in logic.

1. I recently heard a famous logician and philosopher say that "mathematical practice" seems to be today's "holy cow" among certain people. Orientation towards practice-based philosophy of science, of mathematics, of logic, has been a powerful trend in the last few decades. And, at the very least as complement to previous trends, it ought to be welcomed.

2. Readers are likely to be acquainted with the state of academic affairs in relation to logic, and thus to know that the variety of approaches and systems proposed in the field of logic is immensely broader than in mathematics. Just to give an example: a survey of current views about what might be called "basic logic" provides us with many alternatives – traditionalists will still propose first-order logic; more open minded people shall suggest classical logic enriched by modal operators; more radically, an influential group will say IF logic (Hintikka, Sandu); another will insist on epistemic logic (van Benthem).

This would have been extraordinarily surprising for the generation of Frege and Russell, who thought of logic as The One True Logic, and believed it was more solid and strict than mathematics itself. How do we account for that phenomenon?

* This work was supported by the Spanish Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación, under project FFI2009-10024.

It can be understood by placing logic in the context of its associated practices: unicity emerged from systematisation and formalisation of the inferential skeleton of a traditional body of argumentation practices in science and mathematics. It is coherent to think their unicity inherited the homogeneity of a body of practices, far from being *a priori*. Later, when left to itself and to form associations with all kinds of different practices – linguistic, philosophical, computational, etc. –, logic has shown to be highly underdetermined (at least, much more than math).

3. What is a practice-based approach? Simply put: The attempt to ground logical theory on practices of logical inference; e.g. the attempt to solve the vexing problem of the logical constants by arguing from the character of certain (extra-logical) practices and the features of inference and valid argument within them. This means that one argues for the cogency or even necessity of logical features from the problems and goals set for the practice, and considers those logical features as aspects of the proper methodology of the practice.

The main idea in the sequel is simply that classical logic models the inferential structure of traditional practices of representation (and argument) in science and mathematics. I shall speak of mathematico-scientific practices of representation, or even (for brevity) math-scientific practices. But before coming to the details, some preliminaries seems required.

4. It seems necessary to ask twofold: What is a practice? And, how can normative elements emerge from experience within a practice? Here are preliminary answers:

1. A practice can only be discernible – can only exist – when it has an accepted topic and problems (consider e.g. practices of navigation in the Mediterranean by coasting, or in Micronesia by stars, winds, birds;[1] consider the practice of solving practical construction problems in the plane).

2. With an established practice come accepted solutions (examples: to draw a right angle, employ a triangle of sides 3, 4, 5; memorize stars and their positions relative to islands, think of the ship as static and the islands moving towards you).

3. This means that, explicitly of implicitly, conditions of success for problems have been developed. That naturally leads to criteria for success being established, methods being laid out and taught.

4. The initial ingredients for the articulation of normative elements in the practice have been laid. Subsequently may come systematisation, etc., including articulation of criteria as explicit

[1] This famous example of old Micronesian navigation was studied in (Hutchins 1995).

norms.

Note: More than 'normative', perhaps it is reasonable to see logic as being **regulatory**.

5. Attempts to ground logical theory on practices of logical inference (elsewhere) are not very novel. Corcoran in (Corcoran 1973) proposed to understand classical logic as a mathematical model of the inferential structures present "in the practice" of classical mathematics. What may be novel is the meaning we assign to the expression "mathematical practice", the extent to which we mean an idealized form of practice (textbook kind) or else we mean to consider the spectrum of *actual* practices followed or proposed by (real) research mathematicians.

The difference can be crucial, as the case of Corcoran exemplifies. He has a highly specific notion of what mathematical practice is, and his notion is heavily idealized; to put it simply, his math-practice is rigid and also Platonistic. It is easy to show contrary evidence: one can e. g. point to work by individual mathematicians of the first rank that cannot be squared with Corcoran's notion (even if we avoid Brouwer or Martin-Löf, we may consider H. Weyl, W. Lawvere, or T. Gowers). In actual fact, by the mid-20th cent. "mathematical practice" was an expression employed by intuitionists and constructivists above all.

My "mathematical practices" is in the plural, it is not a logician's idealized notion but a historically rich one; it is not rigid but flexible, hence plural.

6. Since my notion of mathematical practice is not rigid, but flexible, it is a problem to explain the incredibly high level of constraint present in mathematical practice as historically given. The mathematics developed by 95% of experts in the last 200 years (notice however the rest 5% includes some of the Abel prizes) can be understood, with some simplification perhaps, to work on the basis of *tertium exclusum* and *reductio ad absurdum*; to be satisfactorily modelled (except perhaps for some traits that one can ignore in good approximation) in deductive reconstructions within the frame of first-order logic; to assume that there is one single structure of the real numbers (an intended model), which in turn can easily be *interpreted* to rest on mathematical Platonism.[2]

Why has that been so? Why such a level of communality? Perhaps it is because the real development of mathematics is an in-time, imperfect expression of ideal mathematics, of Platonic structures that we approximate more and more as we correct methods and results. This is a beautiful perspective that reminds me not only of Gödel, also of French philosopher Lautman, but I find it too difficult to let it go, to be convinced.

[2] Even the celebrated proof of Fermat's last theorem by Wiles, using heavy machinery of algebraic geometry, etc., has been shown to be reducible to axiomatic set theory with minor augmentations; see (McLarty 2010).

There is an alternative: if we forget the 20th century ideology of mathematics for its own sake, math understood as a wholly autonomous discipline,[3] we can bring into the picture the connections between math and physics (let's simplify and forget other disciplines). The argument, presented aphoristically, would be that classical logic is the logic (i.e., the mathematical model of inference structures) of mathematico-scientific representation of phenomena.

7. At a certain level, whose exact boundaries may be difficult to establish, the practices of math-scientific representation share a lot with less specialized practices of representation (like representation of everyday events in the context of legal/practical debates and decisions, or practices of the kind developed by pre-Socratic philosophers). This helps explain how it is possible that Greek authors had already captured the essentials of propositional logic (the Stoics).

At a different level, I believe it is essential for conceptual and historical analysis to insist on the *differences* between Aristotelian (or traditional) logic and the logic of quantifiers. In particular, common representation practices, and even philosophical practices of argument, did not lead to a rich logic of relations and quantifiers. This came from math-scientific practices, and particularly from pure mathematics. De Morgan, around 1860, said: "the algebraist was living in the higher atmosphere of syllogism, the unceasing composition of relations, before it was admitted that such an atmosphere existed." (De Morgan 1966, 241) It was in algebra, he felt, that the general idea of relation emerged, and it was there that "the notions of relation and relation of relation" were first symbolized.

8. How do we get to classical logic from practices of representation? The argument in outline would go this way (please take the sequel as a mere sketch).

Consider **negation**: it seems plausible that all practices of representation, from the common (e.g. involved in navigation or hunting) to the scientific ones (astronomy or geodesy, to name only Ancient sciences), will involve linguistic markers for negation. The relation between proposed models and actually given events (e.g. position of capes in a map vs. actual experience of the capes; orbits of planets vs. their measured positions) will be one of concordance or discordance. This, of course, could and ought to be studied historically, but the goal of representing by means of certain (linguistic or graphic or mental…) representations will make it necessary to speak about *failures* of representation.

However, it seems plausible too that the math-scientific heritage of interest to us has

[3] Perhaps not easy to do, but timely. See for instance the paper: Le retour de Fourier, by J. P. Kahane (2005). URL = http://www.academie-sciences.fr/activite/archive/dossiers/Fourier/Fourier_pdf/Fourier_Kahane.pdf

depended crucially on this: that proposed representations have been subjected to *explicit discursive criticism* (but *if* things are so-and-so, *then*… and this is *not* correct, hence…). We need practices of argumentation, in fact, to have negation inferentially employed in forms like those analysed and systematised by logicians, like those employed by classical scientists and mathematicians. The connection between *modus tollens* and scientific method is too well known to deserve more than a mention.

Take e.g. the Stoics: with practices of representation and argumentation come "assertibles (*axiômata*)", that is to say declarative sentences with a truth-value – at any one time they are either true or false, *tertium exclusum*.[4] And then we have basic inferential schemes like the "five indemonstrables" (modus ponens, modus tollens, hypothetical judgement, etc.), and basic inference rules like the "four *themata*". Thus it seems that one can reconstruct the emergence of **bivalent classical** logic (propositional level) from representation-and-argument practices such as those involved in astronomy or geodesy.

9. It is an interesting historical project to substantiate this concretely; and it seems not only interesting, but also crucial, to inquire the extent to which **realistic assumptions** (of the kind proposed by Plato, Kepler, Gauss) have been influential on the path to these inferential features of the argument practices.

I consider this to be an open question. But notice the issue of realism (or Platonism) is linked with the representation aspect, not primarily with the argumentation. We come back to the topic in connection with intuitionism.

10. I think it will be useful to bring intuitionistic logic (IL) into the discussion. As you all know, intuitionistic propositional logic can be described as classical logic without the Aristotelian law of excluded middle: $(p \vee \neg p)$, but with the law of contradiction $(\neg p \rightarrow (p \rightarrow q))$. In this system, double negation does not imply the truth of the proposition: $(\neg\neg p \rightarrow p)$ is not generally valid; *reductio ad absurdum* is rejected.

Very often, IL is presented as a system codifying the consequence relation in a language that assigns different meanings to the connectives. The new meaning is understood as directly linked, not to truth *simpliciter*, but to actual proof or to actual verification. $(p \vee \neg p)$ would state that either p has been proved or it has been refuted; hence it's easy to see why it fails. $(p \rightarrow q)$ would say that we posses a rule for transforming any proof of p into one of q.

4 See e.g. (Bobzien 2008) in the Stanford Enc. of Phil.

For my purposes, it is important to make clear that this account, although useful for teaching IL (and historically linked with expositions by Heyting), methodologically and philosophically distorts the situation. The intuitionistic laws **are** directly linked with truth and falsity. As R. Cook has argued in detail (see his chapter in Shapiro 2006), one cannot interpret the intuitionist as talking about different connectives.

11. What is then the ground for such a strong difference? I will concentrate on the path followed by Brouwer and Weyl, and it is crucial to pay attention not just to what they proposed, but also to what they rejected – and why. Let us assume, for simplicity of the argument, that Brouwer –like Weyl in 1918, or like he himself up to 1912– has no qualms with natural-number arithmetic. If all there was in math were the natural and real numbers, finite or denumerable structures, and the like, there would be no need for changing logic. But mathematics deals with infinite structures going beyond the denumerable: the structure of real numbers is as central to the enterprise as the naturals.

Now, Brouwer and Weyl reject the usual reconstructions of the structure of real numbers (Dedekind, Cantor, Hilbert) because of their realistic assumptions (Platonism). One cannot treat **R** as a fully determined totality: it's individual elements are not fully determined 'in themselves', for each one of them is an actually infinite system; they are only determined insofar as the mathematician is in a position to offer explicit determinations, which perhaps can be done in denumerably many cases; a fortiori, the "set of all" real numbers cannot be treated as a fully determinate mathematical object – to the indeterminacy of individual reals one must add the fact that, apparently, non-denumerably many of them shall remain indeterminate.

12. The failure of some classical logical laws is a consequence, not of a change in the meaning of the connectives, but of realizing the peculiar methodological and (one might add) metaphysical conditions affecting higher mathematics. There is no continuity between finite structures and higher math, no continuity between representations of real world phenomena and higher math, due to the very peculiar, indeterminate nature of what is usually called "the objects" of higher mathematics.

13. The critic could now say, But the structure of real numbers merely represents the structure of time and of geometric lines! And so many physical theories, from mechanics and electromagnetism to relativity and even quantum physics, have verified empirically the adequacy of such representations! Take for instance QED, which is the most successfully corroborated theory in the whole history of science.

The simple answer is that all of these empirical arguments are insufficient to establish the realistic character of the continuum involved in usual physics. In order to jump to the conclusion that the real number system is fully determinate, we should adopt a fully realistic stance regarding the continuum as a means to represent physical phenomena. Neither all physicists (e.g. Heisenberg, Penrose) nor all mathematicians, nor all philosophers adopt such a stance. Again, there is the option of regarding the continuum (via time or space) as given in intuition, and again this is quite controversial. And, shouldn't logic be a neutral system, independent of such strongly contentious questions?

In the case of Brouwer, precisely it was his intention, from the beginning, to develop pure mathematics as an intellectual system, completely independent from applied math or physics. For a number of reasons, some of them strongly philosophical, he wanted mathematics to remain free from any 'contamination' from the goals and biases of science and technology.[5] I know this kind of proposal is normally disregarded with little or no argument, simply because it is so uncommon. Funny that those who so disregard it will immediately present themselves as advocates of pure mathematics!

14. The whole story of intuitionism and intuitionistic logic, that I have presented too quickly here, should do the job of emphasizing some of the key points. The difference between classical and intuitionistic is not a simple question of change of topic: whatever nuanced change of meaning one may find, its source lies in complex methodological and metaphysical issues. Moreover, it should have become clear that classical logic is intimately tied to math-scientific representation understood in the realistic way that has been characteristic of our tradition (meant is the mainstream scientific tradition, from the Greeks all the way until Einstein and present-day theoretical physicists)[6].

The philosopher of logic does not need to solve all of the vexing problems we have discussed, but in my view he or she should be highly conscious of them. At the very least, one ought to accept that intuitionistic mathematics is a highly relevant theoretical practice based on one sound way of approaching the constellation of problems surrounding the continuum (one among several ways, classical Dedekind-Hilbert approaches included). Thus a practice-based approach to the philosophy of logic is congenial to pluralism, but it should also serve to enhance our awareness of the presuppositions behind different logical systems. By doing this, I hope, we shall be

[5] For details, see the excellent biography (van Dalen 2005). Brouwer's early book *Life, Art, and Mysticism* (1905) is highly relevant for the topic.

[6] Notice than even Kantians, for all the sophistication of their standpoint, leave unaltered the realistic interpretation, now regarding the "phenomenal world" (not the "things in themselves").

contributing to better understanding of what logic is, and what it can be.

REFERENCES

S. Bobzien, "Ancient Logic", *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2008/entries/logic-ancient/>.

L. E. J. Brouwer. *Collected Works,* Vol. I, Amsterdam: North-Holland, 1975.

J. Corcoran, "Gaps between logical theory and mathematical practice", in Bunge M. (Ed.), *Methodological Unity of Science*. Dordrecht: Kluwer. 1973. pp. 23–50.

Dirk van Dalen. *Mystic, geometer, and intuitionist : the life of L.E.J. Brouwer*. Two Vols. Oxford University Press, 2005.

A. De Morgan, *On the Syllogism and other logical writings*, P. Heart (editor), pp. 208–246, Routledge & Kegan Paul, London, 1966.

E. Hutchins, *Cognition in the Wild*. MIT Press, 1995.

C. McLarty. What does it take to prove Fermat's last theorem? Grothendieck and the logic of number theory. *Bull. Symbolic Logic* 16 (2010), no. 3, 359–377.

The Oxford Handbook of Philosophy of Mathematics and Logic, edited by Stewart Shapiro. Oxford University Press, 2005.

Philosophy of Logic, edited by Dale Jacquette. Amsterdam: Elsevier/North Holland, 2007.

H. Weyl, *Das Kontinuum*. Veit & Co., Leipzig, 1918. English translation: *The Continuum*, corrected re-publication, Dover 1994.

LOGICAL CONSTANTHOOD

MARÍA JOSÉ FRÁPOLLI and RAYMUNDO MORADO
Universidad de Granada / Instituto de Investigaciones Filosóficas
mjfrapolli@gmail.com / UNAM, México

ABSTRACT: We defend that the understanding of the role of logic that stems from Frege's Begriffsschrift is compatible with a deeply pragmatist account of the meaning of logical notions. Our concerns are neither the formal expedients labelled as logical constants in artificial calculi nor the natural language terms that are their standard translations. We on a particular kind of concepts, which are trans-linguistic entities that play a specific role in the conceptual systems of rational agents. Our tentative definition of logical constants is the following: Logical constants are higher-order predicables whose arguments are n-adic predicables (n□0), i.e. concepts and propositions. They do not acquire their meaning by referring to extralinguistic entities, neither concepts nor objects, but through the function they perform in representing inferential relations among concepts and propositional contents.
KEY WORDS: logical constants, higher-order predicables, expressivism, pragmatism, Frege, Tarski

1. Introduction

By 'logical constants' we all understand a set of notions that includes at least the following: negation, conjunction, disjunction, the conditional, the universal quantifier and the existential quantifier. To this list, some also add identity, membership, the lambda operator and it might also continue to incorporate non-standard quantifiers, tense, modal and epistemic operators and so on…

What is the philosophical import of asking after the identity criteria for logical constants? This quest is important from the point of view of Philosophy of Language, Philosophy of Mind, or even Philosophy of Mathematics.

1) As a problem for the Philosophy of Language, we ponder what fragments of language, in their usual meaning, may primarily express logical relations. It will not do to simply change the meaning of an expression; in principle we can refer to a logical conjunction even with the help of a disjunctive expression (cfr. Morado 2006).

2) From the point view of the Philosophy of Mind, we are not looking for a class of words but for a group of notions or ways to construct notions, a group of ways to structure our thoughts from a logical point of view. These ways to structure our thoughts allow us to form complex thought out of simpler ones. They are not the only way, but it is a particularly important one for the philosophical question of what is the relation of the structures of our thought to truth and meaning.

Knowing that a mental way of structuring our thoughts is "logical" tells us a lot about the import of that way of structuring to their nature and content.

3) And the identification of a constant as logical has repercussions on our theories of the nature of language, notions or entities we want to identify as computational, mathematical or formal. To classify set membership, value assignment to a registry, lambda conversion or functional application as a logical constant or not, would impact our philosophical understanding of the nature of arithmetic and computability. We can understand better what we are doing when we use certain words, notions or things if we have a principled way to call them logical or not.

The philosophical task of defining logical constants has been approached from syntax, semantics, and there have also been a handful of functional and pragmatic attempts. Invariantist definitions, without exhausting the syntactic perspective, are the kind that has attracted most agreement among logicians. The origin of invariantism is Tarski (1966/1986), with Sher (e.g. 1991, 2003) currently being one of its most active proponents. Gentzen (1935) and, following him, Hacking (1979), Peacocke (1988), and Dummett (1991) have put forward inferential proposals that, being inferential, take a semantic stance on the meaning of these notions. For example, Peacocke (1976:223) introduced a complex approach that included a syntactic criterion, which takes logical constants to be predicative expression formers, and an epistemic criterion, according to which, sequences that satisfy the complex in which logical constants occur can be known a priori. From a pragmatic perspective, Warmbrod (1999) characterizes logical constants by their role in systematizing scientific theories, a characterization very close to the one Ramsey put forward for the case of conditionals (1928:143), but subsequently rejected (1929: 149). Brandom (1994, 2000) offers a normative pragmatic foundation for the meaning of the conditional and negation, and Gómez-Torrente (2002, 2007) suggests some aspects that a pragmatic characterization should cover. In the middle of all this, the possibility of even giving a definition of logical constants that is both, mathematically adequate and philosophically relevant, is still hotly debated (see Gómez-Torrente (2002) and Sher (2003)). However, in spite of all this theoretical activity and exchange of proposals and counter-proposals, we can still claim with Warmbrod (1999:503) that "there is as yet no settled consensus as to what makes a term a logical constant or even as to which terms should be recognized as having this status".

In this setting, the proposal we want to put forward goes along pragmatic lines, and, taking profit from relatively recent developments in the sciences of language, leads to a general understanding of logical constants not as objects but as relations of a certain kind. In fact, our background position is that the answer to the question of what a logical constant is requires a

complete change of scenario, it requires abandoning the path initiated by Tarski and going back to the origins of modern logic, to Frege and Peirce. For the project that Tarski pursued is neither Frege's project, nor the project of a substantial part of contemporary philosophy of logic.

In a first approximation we do not need to distinguish precisely the ontological status of a "logical constant." It is certainly different and important whether we are talking about terms or about concepts. But heuristically we can profit from the double nature of logical constants and the traditional import of logic as a special kind of language and as principles of reason. The ontological, mental and linguistic realms reflect each other closely enough with respect to their logical issues that it is both possible and profitable to disregard their differences in this area. Of course, we will try to call attention to these differences when and if they make a difference for our analysis.

2. Different kinds of logic?

Tarski considers logical truths as the concern of logic. For him (1966:145), logic is "a science, a system of true sentences, and the sentences contain terms denoting certain notions, logical notions". Higginbotham, who also defines logic as "the most general theory of the true and the false" (2000:79), is a contemporary example of this approach. From a first look, Frege seems to be akin to this view in his mature work, since, for example, in 'The thought: a logical inquiry', he claims that "the word "true" indicates the aim of logic as does "beautiful" that of aesthetics or "good" that of ethics" and that it generally "falls to logic to discern the laws of truth" (1918-19/1956:289). Indeed, this is a way of looking at logic that has been very influential, but it is not the only one, and, if we are right, it cannot be the basic one.

Our understanding of logic is based on a different conception, not as the science of truth but as the science of valid inferences, which is also seen in Peirce (Collected Papers, 5.107: 69), in Frege (1879) and in a substantial number of logicians since Gentzen (1935) till today (e.g. Read 2003). It is, of course, undeniable that the very discipline of logic was born from the previous enterprise of applying algebra to the study of natural language (cf. Goldfarb 1979: 357), thanks to the work of mathematicians, such as Jevons, Boole, Peano and Frege, among many others (see Kneale & Kneale, 1962, chapter VI, for a survey). But it is also true that modern logic became an independent science when logicians understood the previous algebraic relations not as relations on sets but as relations on concepts and conceptual contents, i.e. on potential truth-bearers. With respect to this, it should be illuminating to recall how Frege (1882) understood the scope of his *Begriffschrift* and contrasted it with that of Boole's project. Mathematical methods focus on syntactic aspects, and this focus makes it easy to forget that formulae of logical calculi do not stand

for sentences, in the syntactic sense, of natural languages. As Frege himself pointed out (1879:101):

> Since I confined myself for the time being to expressing relations that are independent of the particular characteristics of objects, I was able to use the expression "formula language for pure thought". That it is modelled upon the formula language of arithmetic, as I indicated in the title, has to do with fundamental ideas rather that with the details of execution. Any effort to create an artificial similarity by regarding a concept as the sum of its marks was entirely alien to my thought. The most immediate point of contact between my formula language and that of arithmetic is the way in which letters are employed.

The identification of logic with a branch of mathematics was explicit in Hilbert and his meta-mathematical project, and in the set theoretical approaches to logic due to Zermelo, Fraenkel, Bernays and Gödel. But, as the debate between Hilbert and Frege on the status of axioms made clear, the formal-axiomatic approach to logic was not Frege's project (see for instance Frege's letter to Hilbert, 1.10.1895, in Frege 1980, p. 33).

The contrast between the two ways of understanding logic is enormously relevant for the purpose of defining logical constants, which can thus vary from the task of characterizing some maximally abstract objects to the task of characterizing truth-preserving links among conceptual contents. The two ways determine two independent theoretical enterprises. The first one aims at obtaining maximally abstract formal structures, axiomatically organized, whose axioms are some kind of formal skeletons, 'logical truths', and that deal with sets of highly abstract formal objects, 'logical constants'. This discipline can be considered without risk as a part of mathematics and its objects can be safely defined by mathematical methods. The second enterprise has a different goal: it attempts to represent valid modes of reasoning, with validity rather than truth being its basic notion. It does not purport to identify sets of axioms but rather those transitions between propositional contents that are safe from the point of view of truth preservation. One of the main aims of logic, in this sense, is to offer formal models of the inferential relations produced in particular parcels of natural language. Haack has in mind this discipline when she argues that the "claim of a formal system to be a logic depends […] on its having an interpretation according to which it can be seen as aspiring to embody canons of valid argument" (Haack 1978: 3) .

3. A tentative characterization of logical constants

The characterization of logical constants we propose is [LC]:

[LC]: *Logical constants are higher-order predicables whose arguments are n-adic predicables* (n ≥ 0), *i.e. concepts and propositions. They do not acquire their meaning by*

referring to extralinguistic entities, neither concepts nor objects, but through the function they perform in representing inferential relations among concepts and propositional contents.

[LC] involves a logico-syntactic claim – that logical constants are higher-order predicative expressions; a semantic claim – that they are not referential devices and in particular that they do not name; and a pragmatic claim – that by using them a speaker purports to stress the presence of an inferential link that crucially preserves the truth of the inference at hand. The pragmatic aspects is the essential feature to which the semantic and syntactic aspects have to respond.

Being higher-order n-adic predicables – the syntactic aspect – is what is required if inferential relations have to be relations on (potential) truth bearers. The items connected by inferential links are either full propositions (0-adic predicables) or else n-adic predicables (n>0), in the case of quantified contents.

Familiar intuitions around logical constants, i.e., that they are syncategorematic and topic-neutral expressions, have a syntactic interpretation that is answered for by the syntactic part of [LC]. The items that carry out discourse subject matter are, by definition, the propositions and concepts it includes, i.e. the items that have a truth conditional effect, whereas higher order functions stand at a different level, they play their role regardless of the actual content of the items that are their arguments. Moreover, [LC] incorporates the intuition behind invariantist positions which suggests that logical constants cannot serve to distinguish between individuals: they cannot be used to identify individuals because, being higher-order, individuals do not fall under their scope.

Topic-neutrality also has a semantic interpretation, that is accounted for by the semantic claim of [LC]: logical constants do not contribute to the topic of the discourse in which they occur. And the pragmatic aspect of [LC] gives shape to the intuition that logical constants mark inferential steps in an argument.

REFERENCES

Barwise, J. 1989, *The Situation in Logic*. Stanford, Center for the Study of Language and Information.

Dosen, K. 1989, "Logical Constants as Punctuation Marks". *Notre Dame Journal of Formal Logic* 303), 362-381

Dudman, V. H. 1991, "Interpretations of 'if'- sentences". In Jackson ed.), *Conditionals* (1991), 202-232.

Dummett, M. 1991, *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.

Etchemendy, J. 1983, "The doctrine of logic as form". *Linguistic and Philosophy 6*, 319-334.

Etchemendy, J. 1990, *The Concept of Logical Consequence*. Cambridge, Mass., London, England, Harvard University Press

Frápolli, M. J. 2010, "Naturalizing Logic. The Inference-marker view", in K. Talmont Kaminski and M. Milkowski (eds.), *Beyond Description: Normativity in Naturalised Philosophy*. College Publications, 226-242

Frápolli, M. J. 2012, "¿Qué son las constantes lógicas?", *Crítica. Revista Hispanoamericana de Filosofía*. To appear.

Frápolli, M. J. y S. Assimakopoulos 2012, "Redefining logical constants as inferential markers". *The linguistic Review*. To appear.

Frege, G. 1879, "Conceptual Notation. A formula language of pure thought modelled upon the formula language of arithmetic". In *Gottlob Frege: Conceptual Notation and Related Articles*. Translated and edited with a Biography and Introduction by Terrell Ward Bynum, Oxford, Clarendon Press, 1972, 101-203.

Frege, G. 1882, "Boole's logical Formula-language and my Conceptscript". In *Posthumous Writings*, 47–52.

Frege, G. 1884, *The Foundations of Arithmetic*. Oxford, Basil Blackwell, 1980.

Frege, G. 1892, "On Concept and Object". In Geach, P. and Black, M. (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford, Basil Blackwell, 1952.

Frege, G. 1979, *Posthumous Writings*, trans. Long and White, ed. Hermes, Kambartel, and Kaulbach (Oxford: Basil Blackwell, 1979).

Frege, G. 1980, *Philosophical and Mathematical Correspondence*. University of Chicago Press.

Geach, P. 1962, *Reference and Generality. An Examination of Some Medieval and Modern Theories*. Ithaca and London: Cornell University Press.

Gentzen, G., 1935, "Untersuchungen über das logische Schliessen" ("Investigations into Logical Deduction"). *Mathematisches Zeitschrift* 39: 176–210, 405–431.

Gentzen, G. 1969), *Collected Papers*. Amsterdam, North-Holland

Goldfarb, W. 1979, "Logic in the twenties. The nature of quantifiers". *Journal of Symbolic Logic* 44, 351-368.

Gómez-Torrente, M. 2002, "The problem of logical constants", *Bulletin of Symbolic Logic*, vol. 8, 1-37.

Gómez- Torrente, M. 2007, "Constantes Lógicas". En M. J. Frápolli (ed.): *Filosofía de la Lógica*, Madrid, Tecnos, 179-205.

Gómez-Torrente, M. 2011, "Constante Lógica". En L. Vega y P. Olmos (eds.) 2011, 129-134.

Haack, S. 1978, *Philosophy of Logics*, Cambridge University Press.

Haaparanta, L. (ed.) 2009, *The Development of Modern Logic*. Oxford: Oxford University Press.

Hacking, I., 1979, "What is Logic?" *Journal of Philosophy* 76: 285–319.

Hall, A. 2007, "Do discourse connectives encode concepts or procedures?" *Lingua* 117(1), 149-174.

Higginbotham, J. 2000, "On second-order logic and natural language". In G. Sher and R. Tieszen (eds.), *Between Logic and Intuitionism*. Essays in Honor of Charles Parsons. Cambridge University Press, 79-99.

Hintikka, J. y Sandu, G. 2007, "¿Qué es la lógica?". En M. J. Frápolli (ed.): *Filosofía de la Lógica*, Madrid, Tecnos, 15-54.

Klima, G. 2005, "Syncategoremata". *Elsevier's Encyclopedia of Language and Linguistics*, 2nd. Ed. Edited by Keith Brown. Elsevier: Oxford, vol. 12, 353-356.

Kneale, W. y M. Kneale 1962, *The Development of Logic*. Oxford, Clarendon Press.

Koslow, A. 1992, *A Structuralist Theory of Logic*. Cambridge, Cambridge University Press

Martin-Löf, P. 1983, "On the meanings of the logical constants and the justifications of the logical laws," *Nordic Journal of Philsophical Logic* 1, 11-61.

McGee, V. 2000, "Everything". In Sher and Tieszen (eds.) 2000, 54-78.

Mellor. D. H. 1978), *Foundations: Essays in Philosophy, Logic, Mathematics, and Economics*. London, Routledge and Kegan Paul.

Morado, R. 1987, "Frege, Hempel and Dedekind: Definition of Number and Correferentiality". *Ergo*, vol. I, No. 2, 4556.

Morado, R. 2005, "La representación de las inferencias no-monotónicas". En Ángel Nepomuceno (ed.), *Representación y Logicidad*. Sevilla: Fénix Editora, 37-49.

Morado, R. 2006, "Quine y Fland: ¿Podemos Cambiar el Significado de las Conectivas Lógicas?". In Fernando Martínez Manrique and Luis Miguel Peris-Viñé (eds.), *Actas del V Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España,* Granada, España, 59-66. Reprinted in *Analogía Filosófica*, December, 2007.

Morado, R. 2011. "Lógica/Lógicas". In Vega Reñón and Olmos Gómez 2011, 371-383.

Peacocke, C. 1976, "What Is a Logical Constant?". *Journal of Philosophy* 73: 221–240.

Peacocke, C. 1988, "Understanding Logical Constants". *Proceedings of the British Academy*, 73, 191-200.

Peirce, C. S. 1932, *Collected Papers of Charles Sanders Peirce*. Edited by Charles Hartshorne and Paul Weiss. Volume II. Elements of Logic. Cambridge, Harvard University Press.

Pelletier, F. J. 1999, "A Brief History of Natural Deduction". *History and Philosophy of Logic*, 20 1999, 1-31.

Prior, A. 1976, "What is logic?". *Papers in Logic and Ethics*. Amherst, University of Massachusetts Press.

Ramsey, F. P. 1929, "General propositions and causality". In Mellor 1978, 135-151.

Read, S. 2003, "Logical consequences as truth-preservation". *Logique et Analyse*, 183-184, 479-493.

Sher, G. 1991, *The Bounds of Logic. A Generalized Viewpoint*. Cambridge, Mass. The MIT Press.

Sher, G. 2003, "A characterization of logical constants is possible". *Theoria* 18: 189-97.

Sher, G. and Tieszen, R. eds. 2000, *Between Logic and Intuition. Essays in Honor of Charles Parsons*. Cambridge University Press.

Tarski, A. 1966/1986: "What Are Logical Notions?", *History and Philosophy of Logic*, vol.7, 1986, 143-154.

Toulmin, S. 1958, *The Uses of Argument*. Cambridge University Press.

Vega Reñón, L. and Olmos Gómez, P. eds., *Compendio de Lógica, Argumentación y Retórica*. Madrid, Trotta, 2011. 2012, 2nd edition. Madrid: Editorial Trotta.

Warmbrod, K. 1999: "Logical constants". *Mind*, 108 431: 503-538.

Williams, C. J. F. 1992b: "Towards a Unified Theory of Higher-Level Predication". *Philosophical Quarterly*, 42 169, 449-464

Wittgenstein, L. 1922, *Tractatus logico-philosophicus*. London and New York, Routledge & Kegan Paul.

LAS REPRESENTACIONES Y SU PAPEL OPERATIVO EN LA PRÁCTICA DE LA LÓGICA

VALERIA GIARDINO
Universidad de Sevilla. Spain

"ZETA [*aparte*]: "¡Pragmatismo! ¿Todo el mundo ha perdido el interés por la verdad?"
- Lakatos, *Pruebas y refutaciones*

ABSTRACT: In this article, I will first discuss the difficulties of a *practical turn* in philosophy of mathematics and philosophy of logic. Then, I will present Dutilh-Novaes' thesis according to which formal language is a *technology* and a form of *operative writing* introduced to do logic and overcome the biases of natural reasoning. Finally, I will consider some issues concerning diagrammatic reasoning and discuss its cognitive role. My general aim is to give an idea of how many different areas of research are open once we assume a pragmatic approach to logic and consider the common use of different representations – different technologies in its practice.
KEYWORDS: practical turn, practice-based philosophy of logic, confirmation bias, cognitive tools.

1. De la teoría a la(s) práctica(s)

Hace muchos años se ha introducido el término *practical turn,* 'giro práctico', para describir el desplazamiento del interés filosófico en muchos ámbitos de la individuación de teorías abstractas y generales a la evaluación de lo que los actores hacen practicando y aplicando estas teorías. En este marco, la filosofía deja de dedicarse solo a la definición de lo que cuenta *in vitro* como conocimiento verdadero y justificado y se dirige hacia el estudio *in vivo* de las diferentes prácticas que se encuentran volviendo la mirada a las conductas de actores concretos. La motivación detrás de este cambio está en la crítica a las 'reconstrucciones racionales' típicas de la filosofía tradicional que se han desligado demasiado del mundo; sin embargo, el 'giro práctico' lleva a unas nuevas dificultades, que se pueden resumir en una dificultad general: la vaguedad de la noción de ' práctica'. ¿De qué se habla hablando de práctica? ¿Como se define un 'practicante'? ¿Qué relación hay entre la filosofía y otras investigaciones como la sociología o la historia? Esos son todos temas muy discutidos en la filosofía de la ciencia de las ultimas décadas, a partir de las propuestas de Kuhn y su noción de paradigma – una noción que como la de práctica se ha mostrado peligrosamente ambigua – hasta los enfoques más recientes del naturalismo, que rechazan la

posibilitad de un conocimiento a priori y se identifican con alguna forma de empirismo o de pragmatismo (Kuhn 1967; Giere 1985).

Sin embargo, el 'giro práctico' que ha dominado la filosofía de la ciencia no se ha cumplido todavía más que parcialmente en filosofía de las matemáticas y aún menos en filosofía de la lógica. Como nos cuenta Hersh, en los años 1970 la filosofía de la ciencia ya estaba en su renacimiento, mientras que la filosofía de las matemáticas parecía todavía en su mayoría "un ping-pong fundacionalista" (Hersh, ed. 2005). Una excepción fue *Pruebas y refutaciones*, libro apasionante pero según Hersh entonces virtualmente desconocido. En los años 90, van Bendegem afirma que en matemáticas existe todavía una 'concepción heredada', porque se consideran 'las ciencias exactas' (van Bendegem 1993). La pretensión de identificar las matemáticas con un *estilo de exposición*, comenta Rota, ha obtenido un efecto corrosivo en la forma en que son vistas por los científicos (Rota 1997).

Afortunadamente, en años más recientes, se ha establecido una tendencia muy clara hacia una filosofía de la práctica de las matemáticas, como demuestran algunas publicaciones (véase Mancosu, ed. 2008), muchos congresos y conferencias, y la misma creación de una asociación por la filosofía de la práctica matemática, APMP (véase http://institucional.us.es/apmp/).

Si esta es la situación con la filosofía de la práctica matemática, muy poco se hizo en la dirección de una filosofía de la práctica de la lógica. Grosholz discute como el ideal epistemológico presentado de diversas maneras por Russell, Carnap, y otros para unificar las matemáticas, o las matemáticas y las ciencias, en una teoría formalizada, nace de la demanda de la lógica de proveer un idioma homogéneo como vehículo para la inferencia deductiva (Grosholz 2011). Si esta forma de razonar por sí sola debe transmitir la verdad, entonces todos los términos deben ser estables en su significado, 'iguales' de alguna manera significativa. Según Grosholz, la lógica se mete entonces en problemas a causa de sus propias pretensiones: quiere reemplazar a los lenguajes matemáticos que formaliza. Por lo tanto, ya que los lenguajes lógicos deben ser homogéneos, se ofrecen como sustitutos de (entre otros) los números naturales, llevando al programa de Russell y en un sentido diferente al de Hilbert, y de allí a los teoremas de incompletitud de Gödel. Irónicamente, sostiene Grosholz, la prueba de Gödel es un claro ejemplo de cómo se construye un puente entre formas diferentes de representación, la una útil para hacer referencia, la otra para hacer análisis; Gödel explota una ambigüedad cuidadosamente controlada y fructífera, que multiplica la información disponible y se enfrenta a la demanda de homogeneidad. Sin embargo, es difícil para un lógico admitir que está de hecho "traficando en la heterogeneidad" de las representaciones disponibles.

En mi intervención, voy a considerar un intento muy reciente de 'dar sentido' a la práctica

de lógica y luego a discutir en general el papel cognitivo de las representaciones en lógica, volviendo al tema de cómo en lógica habitualmente se 'trafica' en la heterogeneidad.

2. Dar 'sentido' a la práctica de la lógica: el papel operativo del lenguaje formal

No es controvertido que la práctica de la lógica se ocupe del razonamiento. Pero, ¿entendemos el razonamiento en lógica o el razonamiento de todos los días? En este punto, en la literatura aparece una tensión. Por ejemplo, Shin subraya la importancia de la heterogeneidad del razonamiento afirmando que razonamos válidamente, obteniendo informaciones a través de tipos diferentes de medios de comunicación: diagramas, mapas, olores, sonidos, así como enunciados escritos o pronunciados (Shin 2004). Ella asimila el razonamiento válido al razonamiento de la vida ordinaria que hace uso de tales medios y representaciones. Por otro lado, Corcoran habla de pruebas *gapless*, 'sin huecos', para referirse a pruebas que "describen nuestro razonamiento" (Corcoran, 1969); en este caso todavía, su objetivo no es el razonamiento "normal" o "natural" que por cierto no es riguroso. Sin embargo, reivindica la posibilidad de individuar principios rigurosos en el razonamiento normal y usarlos para definir un sistema deductivo que codifique pruebas rigurosas. Para Corcoran, " práctica deductiva actual" es una práctica *interna* a la lógica, y no pertenece al razonamiento de todos los días, cuyas pruebas, de hecho, son caracterizadas por huecos y omisiones.

Un intento de superar esta tensión lo hizo recientemente Dutilh-Novaes (Dutilh-Novaes 2012). La filosofía de la lógica sigue ocupándose de temas tradicionales como la verdad, la consecuencia, las paradojas, pero no considera los desarrollos más recientes dentro de la lógica misma, concentrándose en la lógica de hace varias décadas, interesada exclusivamente en los fundamentos de las matemáticas (¡el ping-pong fundacionalista de Hersh!). En una palabra, la filosofía de lógica 'actual' no presta atención a la práctica de la lógica 'actual'. Por esta razón, su agenda debería ser ampliada por temas complementarios y no alternativos a los temas tradicionales. Buenas preguntas son las siguientes: ¿cuáles son los mecanismos cognitivos y sociales involucrados en el desarrollo de la lógica como disciplina? ¿Qué es lo que hacemos cuando 'hacemos lógica', y cómo lo hacemos? Una filosofía de la lógica basada en la práctica aspira entonces a 'dar sentido' a las prácticas lógicas; según Dutilh-Novaes, la diferencia principal entre una sociología de la lógica y una filosofía de la lógica basada en la práctica es que la segunda, aunque tenga una dimensión descriptiva y posiblemente sociológica, no puede renunciar a ser de algún modo normativa, es decir debe proveer análisis críticos de los fundamentos conceptuales del trabajo actual en lógica e identificar los posibles problemas conceptuales subyacentes a las prácticas, sugiriendo direcciones

para su perfeccionamiento.

En este marco general, Dutilh-Novaes analiza un problema particular: el papel del lenguaje formal. Considera la investigación en ciencias cognitivas sobre el razonamiento 'en general', con el objetivo de comprender el impacto cognitivo de la introducción de los lenguajes formales para hacer lógica. La idea –en mi opinión muy prometedora– es comprender en qué sentido los lenguajes formales, instrumentos muy poderosos, transforman nuestra manera de hacer lógica. Los resultados empíricos nos muestran lo que ha sido etiquetado como *confirmation bias*, 'sesgo de confirmación'. Con este término general se indica nuestra tendencia irresistible hacia el conservadurismo doxástico (Evans 1989); por lo general, tratamos de confirmar y mantener las creencias que ya tenemos, y somos entonces extremadamente resistentes a la revisión de nuestras creencias previas. Este sesgo de confirmación afecta también a nuestra manera de hacer inferencias: aparentemente nos abstenemos sistemáticamente de sacar conclusiones de premisas que aceptamos si estas conclusiones chocan con nuestras creencias previas. Entonces, nuestras capacidades de razonamiento son minadas por nuestras creencias.

Según Dutilh-Novaes, el lenguaje formal se enfrenta a este sesgo. En primer lugar, es un lenguaje *escrito*, y por eso, tiene (por lo menos) tres papeles diferentes y de algún modo en conflicto: un papel expresivo, un papel icónico y un papel operativo. El tercero pertenece al cálculo y define su 'poder computacional'. Además, está estrechamente relacionado con el segundo y se conecta con la noción de 'escritura operativa' (*operative writing*) introducida por Krämer (2003). En el caso de los lenguajes formales, esta noción concierne a la dimensión *paper-and-pencil*, 'papel-y-lápiz'. El lenguaje formal, en su papel operativo, hace posible un proceso de 'de-semantificación': es decir, quien razona a través del lenguaje formal deja de ser guiado en su inferencias por intuiciones o creencias que se activan casi automáticamente, porque su única tarea es la aplicación 'mecánica' de unas reglas definidas dentro del formalismo; en consecuencia, el lenguaje formal, operativamente, le permite economizar sus recursos cognitivos. Haciendo inferencias por el medio del lenguaje formal, el lógico sabe *cómo* llegar a una conclusión sin necesariamente saber *porqué*. Además, una vez que el lógico tenga el lenguaje formal a su disposición, puede también iterar sus operaciones sin límite, diferentemente del agente guiado por las intuiciones y el significado que no puede abandonar el territorio cognitivo donde le es posible 'intuir'.

La tesis de Dutilh-Novaes es que los lenguajes formales como otros lenguajes operativos en general constituyen un tipo particular de *tecnología*, cuyo uso permite a un agente humano "ejecutar un programa totalmente diferente", por fin remplazando el programa que se ejecuta normalmente

razonando y que lleva al sesgo de confirmación. Como otras tecnologías, los lenguajes formales han sido introducidos y desarrollados en vista de una particular función: sirven como contrapeso a algunos de nuestros patrones cognitivos más espontáneos, activando procesos cognitivos que podemos ejecutar pero que no se activarían bajo circunstancias habituales.

Pues quizá los lenguajes formales hayan por fin abierto nuevas posibilidades no previstas. Dutilh-Novaes lo llama *surprise factor*, 'factor sorpresa', muy típico en el desarrollo de una tecnología. Puede ocurrir entonces que los lenguajes formales, adoptados para compensar algunas limitaciones del nuestro razonamiento 'normal', han acabado por tener una 'vida propia' y crear nuevas prácticas que no eran el objetivo inicial de su introducción. Por ejemplo, podemos imaginar que su uso ha llevado por fin a la concepción de lógica como idioma homogéneo mencionada por Grosholz; esta interpretación es coherente con la idea que los lenguajes formales nacen como lenguajes 'de-semantificados'.

3. El papel cognitivo (operativo) de las (diferentes) representaciones en lógica

Comparto las conclusiones de Dutilh-Novaes, pero aquí voy a proponer un camino diferente: para evaluar el papel operativo de las representaciones, intento mirar directamente a su 'operatividad'.

Consideremos el razonamiento por diagramas. Como los lenguajes formales, los diagramas también se pueden considerar formas de escritura operativa. Peirce define el 'diagrama-icono' (recordamos que el papel operativo se relaciona estrechamente con el papel icónico) como no estático sino dinámico: el diagrama-icono determina un estado de actividad en su intérprete que se mezcla con la curiosidad (Peirce 1906). Es decir, un diagrama es siempre *interpretado* dentro de un contexto de uso particular y además expresa una intención subyacente, siendo concebido en vista de algún objetivo particular. En este marco, un diagrama también es una tecnología. Pero, si es una tecnología, ¿como funciona?

Imaginemos el diagrama de un círculo, que se ha usado en lógica, en geometría euclidiana, o en geometría cartesiana. ¿Es en todo caso 'el mismo' círculo? ¿Tiene siempre el mismo papel? En un cierto sentido sí, porque en todo caso tiene un papel operativo. Sin embargo, también no, porque no existe un conjunto específico de reglas que se puedan fijar de una vez por todas para extraer su significado.

De hecho, para aprender a usar un diagrama –por ejemplo un círculo– no nos referimos a sus propiedades visuales o a reglas establecidas previamente, sino sobre todo al sistema de 'permisos' o licencias que hemos aceptado para *manipularlo* con propósito inferencial dentro de un sistema de

representación: aprender a usar un sistema particular de diagramas para hacer inferencias es aprender una *práctica de manipulación*. El diagrama se ofrece como un laboratorio donde el lógico hace operaciones, planos y experimentos con el objetivo de encontrar soluciones y razones para sus soluciones. Los agentes, una vez que hayan aprendido la práctica, comprenden y reproducen las representaciones en una manera no mecánica y 'sin daños'. Por ejemplo, quien esté ya dentro de la práctica de la geometría euclidiana sabe perfectamente, sin haberlo aprendido de antemano por medio de una regla explícita, que una vez que ha construido la figura, puede rotarla o moverla en el plano, pero no puede por ejemplo estirarla, a diferencia de lo que podría hacer dentro de otras prácticas.

Mi hipótesis es que en lógica como en matemáticas y más en general en una práctica que implica el uso de representaciones no lingüísticas, activamos una forma de 'imaginación manipuladora', es decir una imaginación que resulta particularmente efectiva porque estamos ya familiarizados con ella, ya que en cierta medida reproduce procedimientos similares a la manipulación de objetos concretos derivada de nuestra experiencia perceptiva. Sin embargo, no todas las manipulaciones posibles son permitidas dentro de una práctica especifica: las manipulaciones que se imaginan activamente en el diagrama son controladas por la interpretación y la práctica compartida. Además, un aspecto importante de los diagramas es su ambigüedad: un diagrama puede ser 'leído' (interpretado) de diferentes maneras, y sólo la práctica y los procedimientos compartidos fijan su significado. La ambigüedad, como sugería Grosholz, no es tanto una desventaja en principio, sino un punto de fuerza del razonamiento, que puede ser promotor de inferencias (Giardino, de próxima aparición 2012).

Mi posición concuerda con el marco propuesto por Dutilh-Novaes: como el lenguaje formal, los diagramas también se pueden entender como formas de escritura operativa que ofrecen una tecnología para promover inferencias y economizar nuestros recursos cognitivos. Aún así, considero que cada una de estas formas de escritura se basa en mecanismos que pueden ser diferentes y que merecen ser investigados en profundidad. Dutilh-Novaes se da cuenta que el uso de lenguajes diferentes para enseñar la lógica, por ejemplo sistemas de representación por enunciados o por diagramas, tiene un impacto significativo sobre el proceso de aprendizaje (Stenning 2002). En el caso del lenguaje formal, podemos hablar de de-semantificación, mientras en el caso de instrumentos más icónicos, aunque igualmente operativos, podemos hacer la hipótesis de una activación de una forma de imaginación manipuladora. Vemos entonces que sólo poniendo atención a las representaciones comunes, numerosas áreas de investigación se abren. Es probable que otras áreas se abrirían considerando por ejemplo las prácticas actuales de los lógicos, que es otra

cuestión empírica y no trivial. Se muestra así el interés de abandonar un enfoque puramente abstracto en el análisis de la lógica, asumiendo un enfoque que incluya la pragmática. Como señalan Grosholz o Shin, nuestro razonamiento está determinado por la riqueza de nuestros recursos cognitivos y de las representaciones que introducimos. Es hora de que la filosofía de la lógica (¡de la práctica de la lógica!) dé sentido a la naturaleza del razonamiento: su heterogeneidad.

REFERENCIAS

Corcoran, J. 1969, 'Three Logical Theories', *Philosophy of Science* 36, 153-177.

Dutilh-Novaes, C. 2012, 'Towards a practice-based philosophy of logic: formal languages as a case study', in *Philosophia Scientiae* 16(1), 71-102.

Evans, J.St.B.T. 1989, *Bias in human reasoning: Causes and consequences*, London, Erlbaum Associates.

Giardino, V. de próxima aparición 2012, 'A Practice-based Approach to Diagrams', en Moktefi, A. and Shin, S.-J., ed., *Visual Reasoning with Diagrams, Studies in Universal Logic*, Birkhauser, Springer.

Giere, R. N. 1985, 'Philosophy of Science Naturalized', *Philosophy of Science* 52, 331-356.

Grosholz, E. 2011, 'Logic, Mathematics, Heterogeneity', en Cellucci, C., Grosholz, E. and Ippoliti, E., ed., *Logic and Knowledge*, Cambridge Scholar Publishing, Newcastle Upon Tyne, 411-426.

Hersh, R., ed. 2005, *18 Unconventional Essays on the Nature of Mathematics*, Berlin, Springer-Verlag.

Krämer, S. 2003, 'Writing, notational iconicity, calculus: On writing as a cultural technique', *Modern Languages Notes – German Issue*, 118(3), 518-537.

Kuhn, T. 1962, *The Structure of Scientific Revolutions,* Chicago, University of Chicago Press.

Mancosu, P., ed. 2008, *The Philosophy of Mathematical Practice*, Oxford University Press.

Peirce, C. S. 1906, 'Prolegomena To an Apology For Pragmaticism', *The Monist* 16(4), 492 -546.

Rota, G. C. 1997, 'The Phenomenology of Mathematical Proof', en Kamamori, A., ed., *Proof and Progress in Mathematics*, Dordrecht, Kluwer, 183-196.

Shin, S.-J. 2004, 'Heterogeneous Reasoning and Its Logic', *The Bulletin of Symbolic Logic* 10(1), 86-106.

Stenning, K. 2002, *Seeing Reason*, Oxford, Oxford University Press.

van Bendegem, J. P. 1993, 'Real-life Mathematics versus Ideal Mathematics: the Ugly Truth', en Krabbe, E. C. W., Dalitz, R. J., Smit P. A., ed., *Empirical Logic and Public Debate, Essays in Honour of Else M. Barth*, Lanham, Rowman & Littlefield, 172-182.

PRACTICE-BASED VERSUS PRAGMATIC APPROACHES TO LOGICAL CONSTANTS*

CONCHA MARTÍNEZ-VIDAL
University of Santiago de Compostela. Spain
mconcepcion.martinez@usc.es

ABSTRACT: There is a sense in which practice-based approaches could be considered pragmatic approaches; namely and trivially, practice-based approaches take practices into account. But, given some existing characterizations of practice-based approaches, not all pragmatic approaches are practice-based. In the following, I intend to illustrate the problem by showing that Frapolli's pragmatic approach to the characterization of logical constants is not practice-based. I will start by typifying both practice-based and pragmatic approaches, and continue by applying those characterizations to Frapolli's proposal, and to the so-called view of logic as model (Corcoran 1973; Shapiro 2001;Martínez-Vidal 2004). My contentions will be: 1) that the view of logic as model is a practiced-based approach, while Frapolli's pragmatic approach is not; 2) that pragmatic approaches cannot be trivially identified with practice-based ones.
KEYWORDS: pragmatic, practice-based, the view of logic as model, logical constants.

1. Practice-based and pragmatic approaches to the philosophy of logic: a characterization

1.1. Practice-based Approaches.

One important recent contribution to the development of the practice-based approach to the philosophy of logic was the Practice-based Philosophy of Logic and Mathematics Workshop celebrated in Amsterdam in 2009. Before, there were different results in what may be called practiced-based philosophy of mathematics, and some methodologies in the philosophy of logic that, I take, should be taken to be practice-based, for instance the view of logic as model approach advocated by Corcoran (Corcoran 1973). There were also developments in the philosophy of language and logic according to which our linguistic practices are fundamental to meaning in general, and to the meaning of logical constants in particular, given that "the meaning or propositional content of an expression or attitude is determined by the role it plays in a person's

* This work was supported by the Spanish Ministerio de Economía y Competitividad, Secretaría de Estado de Investigación, Desarrollo e Innovación under project FFI2009-08828 and by the Portuguese Science Foundation (PTDC/FIL-FIL/109882/2009, FCT)

language or in her cognition."[7] Frapolli's proposal (this volume) is a case in point. One aim of this contribution is to elucidate to what extent all these pre-existent strategies can be considered practice-based.

Fundamental to our purposes will be to count on an appropriate characterization of practice-based approaches. According to Dulith Novaes practice-based approaches to logical constants are to comply with the following requisites (Dulith Novaes 2009):[8] i) they should take actual logical practices as starting point rather than idealized notions; ii) they must gather empirical data on how research is actually conduced; iii) they should clarify underlying assumptions, raise pertinent questions, draw philosophically important conclusions from technical results – 'make sense' of logical practices; iv) in relation to those practices, they should consider both the individual (cognitive) and the social level. The latter is not to be identified with the sociology of logic which is essentially descriptive: practice-based accounts have a descriptive dimension, but they are also prescriptive. Actually, she contends that one major challenge for practiced-based approaches is how to deal with normativity—with "how things ought to be, […] given how things actually are"—to suggest that the solution is to apply the methodology of reflective equilibrium. I take it that what is relevant for an approach to be practice-based is that is has to deal with this problem, though it should not be necessary to subscribe to Dulith Novaes favorite solution; v) they are to offer critical analysis of the conceptual foundations of actual work being done in logic (clarifying underlying assumptions), possibly to identify conceptual problems underlying the practices and suggest directions for improvement; vi) It need not replace traditional philosophy of logic.

The gathering of data presents one difficulty the author herself acknowledges; namely,

> "[h]ow can we gather data about actual practices in a methodologically robust manner? Mere anecdotic evidence is not sufficient […] For the individual level of logic as a cognitive enterprise, a promising approach seems to be to take into account findings from cognitive science. […] For the collective level of logic as a social enterprise, data-gathering seems a delicate matter. Serious sociological methodology would have to be employed. (Surveys?)" (Dulith Novaes 2009)

The problem is that this emphasis on gathering empirical data without further requirement can misguide our research: we do not want to do either psychology or sociology of logic, as the authoress herself acknowledges. For instance, Tarski intended to characterize the preexistent notions of truth and logical consequence, but, making a survey on what people understood by those

[7] Whiting 2009.

[8] In what follows, I paraphrase, almost literally, her proposal: http://www.illc.uva.nl/pplm/ uploaded_files/2What_is_practice-based_philosophy_of_logic_line.pdf

notions was not the kind of data that had to be gathered. Hence, the role of practice, and the kind of methodology for the gathering of data need to be qualified. To clarify what this gathering of empirical data would amount to, I have considered Dulith Novaes examination of the role of formal languages in logic, the kind of analysis performed by other participants at the Amsterdam Workshop, [9] and the kind of scrutiny practice-based approaches in the philosophy of mathematics have actually done (Mancosu 2008). To illustrate the problem of normativity, in section three below I will rely on Shapiro's approach to the problem (Shapiro 2001).

And this is what different philosophers who see themselves as applying a practiced based methodology take to be relevant to study logical or mathematical practice:

- Case analysis both historical and contemporary
- Study of the different processes involved in getting to know logic; not only proof, but also the cognitive processes involved in proving. In general, all processes and methods involved in the causal chain that allows a particular logician or community of logicians to come to know. Thus, it could study the role of visual reasoning, the changes brought about by the use of logic in non-traditional fields (computer science, for instance), the use of computers, etc.
- Historical analysis of the growth of logical knowledge in terms of the way practices have changed.[10]
- Characterizing explanation and understanding in logic;
- Addressing classical problems in the philosophy of logic (or mathematics) from this new perspective.

1.2 Pragmatic Approaches

My characterization of pragmatic approaches to logical constants elaborates on the proposals by Warmbrod (Warmbrod 1999), Gómez-Torrente (Gómez-Torrente 2002, 2007), and MacFarlane (MacFarlane 2009). Pragmatic approaches are usually characterized as opposed to analytic approaches (MacFarlane 2009), that is to say, to those that start with a property that logical constants are to comply with: being formal, topic neutral, etc. On the contrary, pragmatic approaches "… start with a *job description* for logic and identify the constants as the expressions that are necessary to do that job. For example, we might start with the idea that the job of logic is to

[9] Practice-based approaches to logic and mathematics, Amsterdam 2009.

[10] I am paraphrasing Mancosu when he accounts (Mancosu 2008) for what the new tradition involving Lakatos, Kitcher, Tymozcko aims at.

serve as a "framework for the deductive systematization of scientific theories" (Warmbrod 1999, 516), or to characterize (classical) mathematical structures and represent (classical) mathematical reasoning (Shapiro 1991), or to characterize the logic underlying given inferential practice (Corcoran 1973).

There is an ongoing debate between Warmbrod and Gómez-Torrente as to whether pragmatic characterizations are complete or incomplete. Warmbrod contends that pragmatic demarcations are guided by what he calls a "requirement of minimalism":

> "…logical theory should be as simple, as modest in its assumptions, and as flexible as possible […] In practice, the minimalist constraint dictates that the set of terms recognized as logical constants should be as small as possible." (Warmbrod 1999, 521)

But Gómez-Torrente rejects this minimality requisite:

> "Warmbrod has misidentified the source of the difficulties facing typical characterizations of logical constancy. The problem is not that a principle-based characterization is impossible *tout court*. In fact, I think […] that a logical constant may be just an expression which satisfies some appropriate basically pragmatic principles. The complexity of the principles, and also their somewhat vague nature, does much work to avoid any easy refutation of a proposal of this sort —either of the claim that it provides necessary conditions for an expression to be logical, or of the claim that it provides sufficient conditions. The problem with typical characterizations is that they have no place for that sort of complexity, as they try to unearth necessary and sufficient *mathematical*, *semantic* and *epistemic* properties for an expression to be logical; it is a characterization of this sort that seems hopeless." (Gömez-Torrente 2002, 34)

We agree with Gómez-Torrente that a complete definition is possible if it considers pragmatic requisites. The problem, of course, is whether these pragmatic characterizations lack the philosophical interest analytical characterizations claim to have (Gómez-Torrente 2002, 35-36). Also, MacFarlane (MacFarlane 2009) advocates some additional features:

- Pragmatic approaches tend to be holistic while analytic approaches tend to be local.
- On pragmatic demarcations what is logical might depend on the current state of scientific and mathematical theory (Warmbrod 1999, 533), while "[o]n an analytical demarcation, whether particular resources are logical depends only on whether they have the favored property." (MacFarlane 2009)

The point is what the relation is between these two kinds of methodologies: are all pragmatic demarcations practice-based? Are all practice-based demarcations pragmatic?

2. Frápolli's approach: a pragmatic approach in Gómez-Torrente's sense

Frápolli's argument (Frápolli forthcoming) to contend that only a certain kind of negation, the conditional, the biconditional, the universal quantifier with conditional, and non-standard quantifiers with the conditional are *genuine* logical constants, goes as follows:

1) In order to characterize logical constants, we need to start by clarifying what 'logic' means. Her preferred notion of logic is what she calls Frege logic (F-logic); it is the conception of logic as the science that studies inference, inferential transitions.
2) She also claims we need to be aware of the difference between i) understanding the meaning of those expressions in natural language that are the counterparts of logical constants in logical calculi, and ii) understanding how those formal counterparts work in their calculi. [11] This is central because in order to clarify what a logical constant is, we need to understand the meaning of those expressions in natural language.
3) Now, the disciplines that study the meaning of expressions and concepts (at one level or other) are linguistics and the philosophy of language. Linguists tell us that logical constants work as 'discursive markers', though not all discursive markers are logical constants. Markers do not add content; they codify rules, while the rest of expressions in natural language codify concepts.
4) She also claims that this conception of logical expressions as markers suits properly Frege's conception of logic, and that her approach works independently of which theory of meaning is chosen. In other words, she contends that given what she calls a Fregean conception of logic, and observing the practice of speakers results in subscribing to the view that the job of logical constants is to be inferential markers.
5) She proposes (CL), a necessary and sufficient conditions for a logical expression to be a logical expression. (CL) establishes syntactic, semantic and pragmatic conditions for an expression to be a logical constant. (CL) asserts that the job of logical constants is to codify inferential relations between concepts and between propositional contents.
6) Only the expressions mentioned above are genuine logical constants.
7) More comprehensive sets of expressions will be logical, if different functions for them are considered.

Frápolli's characterization is a pragmatic one—according to the characterization of pragmatic approaches given above—because it clearly spells a job for logical constants in order to demarcate that set, yet, the job is not the starting point. Her starting point is a particular conception of logic

[11] I am paraphrasing Frápolli (Frápolli forthcoming, section 2).

that leads her to choose a certain theory of meaning for logical constants, one in which the pragmatic element plays an essential role. As a result, she offers a necessary and sufficient condition for genuine logical constants to be such. Given that her principled characterization takes into account pragmatics, her proposal would count as evidence to Gómez-Torrente's point that in order to give necessary and sufficient conditions for a logical constant to be such, pragmatics is to come in; but then, she seems to be claiming that her characterization captures something that is philosophically relevant when she claims there is a set of *genuine* logical constants. Nevertheless, she also claims that a deeper knowledge of logic is attained if it is accepted that different sets of expressions can be logically interesting depending on the specific functions we ask them to comply with.[12] It seems to me there is a certain tension between these two claims; arguing for the existence of 'genuine logical constants' seems to point at a principled account, while her acknowledging that different set of expressions can be logically interesting pulls towards a non-principled one.

Frápolli's approach would not be a practice-based approach because it starts with an idealized notion, and she emphasizes the existence of a *genuine* set of logical expressions even if it pays attention to our practices. Her approach would be a principled one. We are to pay attention to practices because they are fundamental to her favorite conception of logic. Her methodology does not seem to involve gathering empirical data on how research is actually conduced; on the contrary, her approach sets constraints from above, starting with a certain conception of logic. Given her approach, the philosopher of logic tells the logician which set of expressions should be considered *genuinely* logical.

3. The view of logic as model

The basic idea of the view of logic as model (Corcoran 1973) is that before the first logical system was developed by Aristotle, there existed a large amount of proofs. The logician's task consisted (and consists) in developing models to account for the underlying logic for mathematical practice, that is to say, to account for what a proof is.[13] Mathematical practice is supposed to play the role scientific facts play in the case of natural science. In the case of natural sciences, once a model is built, it is tested by checking how well it works in explaining and predicting scientific facts. The analogy goes that logical systems are mathematical models used to explain and 'predict' mathematical practice.

This methodology can be seen as practice-based since i) it takes as starting point actual

[12] Frápolli forthcoming my translation.

[13] Other authors share with Corcoran the view of logic as model, though their exact formulations may not be identical: Burgess 1992, Shapiro 1998, etc.

logical practices rather than an idealized notion of logic or logical constant; ii) it gathers empirical data; iii) it tries to clarify which is the underlying logic of a given practice; iv) it has a descriptive dimension, but it has the kind of problem Dulith Novaes signaled: "how things ought to be, […] given how things actually are". Shapiro (Shapiro 2001) identified the problem: when the logician is modeling the mathematician's practice and the logician realizes the mathematician is applying a rule the logician identifies as invalid, if the source of objectivity is mathematical practice, there is no reason whatsoever to force the mathematician, who never comes to a false statement by applying what the logician takes to be an incorrect rule, to abandon her practice. Notice that in other fields where mathematical models are used it is absurd to think that one could use the model to maintain, for instance, that the stars fail to follow the model. The problem with the view of logic as model is that since what is to be modeled is the mathematician's practice, the logician cannot correct the mathematician (practice goes first), but, intuitively, we want an account of logic that allows the logician to correct the mathematician.[14]

Back to our methodological worries, the view of logic as model can also be considered a pragmatic approach, according to the MacFarlane-Gómez-Torrente characterization given above: it establishes a job to be done, namely, to characterize the logic underlying a given mathematical practice. Once the logic is characterized, which expressions are to be considered as logical will be clear, namely, the ones in the underlying logic.

Conclusion

Pragmatic based methodologies and practiced-based methodologies cannot be trivially identified. It is not clear that all pragmatic approaches are practice-based as our assessment of Frápolli's proposal shows.

As to the problem of logical constanthood, it seems clear that, given this methodology, the problem becomes that of characterizing first the logic underlying a given practice, and then to establish which is the set of logical constants in that logic.

REFERENCES

Corcoran J. 1973, "Gaps Between Logical Theory and Mathematical Practice" in Bunge (Ed), *The Methodological Unity of Science*, Reidel: Boston, USA, 23-50.

Dulith Novaes C. 2009, "What is practice-based philosophy of logic? A case-study: uses of formal languages in logic", Workshop Practice-based Philosophy of Logic and Mathematics, Amsterdam 2009, http://www.illc.uva.nl/pplm/index.php%3Fpage=_.html

Frápolli MJ. Forthcoming, , "¿Qué son las constantes lógicas?".

[14] For complete discussion of this issue see Shapiro 2001; Martínez-Vidal 2001, 2004.

Gómez-Torrente M. 2002, "The problem of logical constants", Bulletin of Symbolic Logic, vol. 8, 1-37

Gómez- Torrente, M. (2007), "Constantes Lógicas". En M. J. Frápolli (ed.): Filosofía de la Lógica, Madrid, Tecnos, 179-205.

Kitcher P. 1984, *The Nature of Mathematical Knowledge*, OUP: New York, USA.

MacFarlane, John, "Logical Constants", *The Stanford Encyclopedia of Philosophy (Fall 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2009/entries/logical-constants/>.

Mancosu P. (Ed.) 2008, *The Philosophy of Mathematical Practice*, OUP: USA.

Martínez Vidal C. 2001, "The view of Logic as model: Comments on Stewart Shapiro`s paper", AGORA, 20, 1, 175- 179.

Martínez Vidal C. 2004, "´The View of Logic as Model and the Objectivity of Logic" Reti, Saperi, Linguaggi Revista del Dipartamento di Scienze Cognitive e della Formazione dell`Università di Messina, Volume 1, 35-74.

Shapiro S. 1991, *Foundations without Foundationalism. A Case for Second-order Logic*, Clarendon Press : Oxford.

Shapiro, S. 2001, "Modeling and Normativity: How much revisionism can we tolerate?", *AGORA*, Vol. 20, Nº 1, 2001 , 159-173.

Whiting D., 2009 "Conceptual Role Semantics", Internet Encyclopedia of Philosophy, http://www.iep.utm.edu/conc-rol/

DEDUCTIVE PRACTICE AND METHODOLOGICAL THEORY

JOSÉ M. SAGÜILLO
University of Santiago, Spain
josemiguel.saguillo@usc.es

ABSTRACT: The issue of this paper is whether mathematical logic is descriptive of the underlying logic exhibited in mathematical practice or whether it is rather normative in indicating how such practice should be conducted. Once the meaning of 'normativity' in the present context is clarified as pertaining to pragmatics, the question arises where to locate normativity, whether in mathematical practice, in logical theory, or in both. The discussion is framed under the view of Logic as model (Corcoran 1973), the view that mathematical logic is about the underlying logic found in the prior practice of proof in real mathematics, and what I call Shapiro's dilemma (2001), the difficulty in articulating how the view of logic as model could illuminate in what sense—if any—a given logical theory (a mathematical model of logic) can be said to be normative or can in fact be used to discriminate good from bad deductive practice.
KEY WORDS: Logic, practice, proof, description, model-artifact, normativity

Introduction: Shapiro's dilemma

There is a fruitful analogy between models of the natural sciences, let us say, as in the planetary model of subatomic particles, and models of logic. If we take logic to be about reasoning—the analogy goes—then mathematical logic is just an applied branch of mathematics whose aim is to build "artifacts" which model good and bad reasoning. For example, a formal system of derivation codifies in a formal language features of good deductive practice usually found in a natural language. Likewise, a model-theoretic semantics codifies features of a conception of validity and invalidity, which is also found in practice and which, in turn, sustains the soundness and unsoundness of derivational moves. Mathematical logic produces such model-artifacts with the material provided by string theory and set-theory. So far so good.

In current literature one often finds locutions, such as "Logic is normative" to the effect that Logic is not just descriptive of logical phenomena underlying mathematical practice but also that it carries the status of *canon* of good reasoning or better, the status of "norm" dictating how *we ought* to reasoning. Under the assumption that Logic is normative Shapiro (2001) claims that the view of logic as model does not illuminate in what sense—if any—a given logical theory (a mathematical model of logic) can be said to be normative or can in fact be used to discriminate good from bad

deductive practice. Shapiro, following Corcoran (1973), correctly indicates the inherent existence of gaps—either of excess or defect—between the model and that which the model is a model of. He argues that we may be at lost trying to discriminate between genuine flaws in the model from [objectively] bad deductive practices. The question arises—Shapiro concludes—on what grounds a logician supplying such a model can accurately ("legitimately") point out mistakes in the deductive practice of the working mathematician?

One important issue in assessing the analogy is the fact that in physics, for example, if a prediction made according to the model is disconfirmed scientists do not *blame* reality for the failure but they try to identify a gap in the descriptive power of the model. Shapiro, on the other hand, would like to use the view of logic as model to assess how much normativity a logician can impose on defective, or perhaps divergent, practice from the viewpoint of the correctness of the model. Shapiro's discussion is disappointing—as he himself concedes—in that no definite answer to his dilemma is provided. I explore a tentative—perhaps even more disappointing—response to his dilemma by exploiting the bounds of the analogy between *the use* of models in natural sciences and *the use* of models of logical phenomena which is sustained on a certain metaphysical conception of laws, whether laws of nature or laws of thought. I argue that there is no normativity in a model of reasoning *per se* although I grant that a certain use of *any* model can be "normative" in practice as it happens, for example, in the context of teaching and training in a lab or in a logic course. Leaving aside these cases it is shown that the view of logic as model cannot help with the view of logic as normative. Thus, a more definite negative answer to Shapiro's dilemma is provided. I postpone this discussion to the second part of the paper. Let me first juxtapose several views about the *locus* of the "normative mode" in order to propose a way it can be coherently articulated.

Exchanges and "pulls" between mathematical practice and logical theory

Undoubtedly the relation between mathematical practice and logical theory is complex and has been the subject of philosophical disputes. One controversial issue is the location of normativity, whether in the prior realm of mathematical practice, or whether in the realm of models of reasoning, or in both.

For example, Solomon Feferman (1964) supports the priority of mathematical practice over logical theory when studying correct reasoning:

> "Logical thinking in mathematics can be *learned only by observation and experience. In fact, the ability to reason* correctly and to understand correct reasoning is itself a prerequisite to the study of formal logic".

Feferman seems to imply that the ability to think and to reason correctly is experientially learned in mathematical practice. I have suggested in a previous paper (Sagüillo 2006) that the only kind of modality involved in a proof is epistemic and not normative, at least in the usual sense of 'normative'. A proof involves an intrinsic epistemic necessity ruled by the mind of the thinker for whom a piece of reasoning is actually a proof. Epistemic necessity should be contrasted with the alleged *extrinsic* normative necessity we often encounter in a social system governed by rules or norms. This second kind of necessity appears in the realm of the practice of proving arising perhaps in a particular context of teaching or instruction and its part of the curriculum of pragmatic studies. More precisely, understanding logical reasoning in a proof is certainly not mandatory as a traffic norm is, nor it can be imposed on anyone whether in a real proof or in a reconstructive situation guided by a set of rules for derivation. Mandatory devices, such as rules and norms, do their job but they *really work* only when people interiorized the underlying *good* reasons to follow the given rule or the given norm. Of course this is not to deny that we often do not obey the rule or the norm even though we fully understand its meaning and purpose in the same way we often make performative mistakes of reasoning in spite of being—at least potentially—competent thinkers. Grasping the cogency of a chain of reasoning is something that requires the free and alert activity of the mind. In short, a thinker is rationally convinced or it is not convinced at all.

On the other hand, Patrick Suppes (1957/99: 128-129) illustrates a normative use of logical theory when discussing the relation between vague deductive practice and formal proof:

> "The very vagueness of the criteria governing informal proofs is a primary justification for a precise definition of a formal proof […] In cases of controversy or doubt concerning the validity of an informal proof it is exceedingly useful to have available a clear, exact standard to which appeal may be made".

I already argue that the kind of normative use of a model suggested by Suppes is not sufficient to prompt understanding and thus cogency of a given step of reasoning in the mind of the thinker. Moreover, cogency cannot be forced upon by rules or norms, or by any form of appealing to authority, imitation or force. Cogency can be "detonated" instead with *instruction*—as Feferman indicates—which seems to be prior to the study of a mathematical system of derivation.

We could also find qualified opinions to the effect that the difference between an informal and a formal proof is perhaps just a matter of degree. Corcoran (1992: 39) identifies historically this kind of scenario:

> "As far as we know rudimentary axiomatizations of parts of the body of then-known mathematics were attempted more than a century before Euclid who

> flourished around 300 B.C. But for centuries before the rudimentary axiomatizations, Greek mathematical activity embodied several practices that can be seen in retrospect to be essential to the axiomatic method. It should be no surprise that the components of the axiomatic method had been in use before being combined into a method".

Among the previously available practices, proof procedures involving deductions and demonstrations were fundamental. Moreover, the abstraction required for the study and articulation of the axiomatic method has as a precondition the existence of an already available stock of concrete informal proofs. If this is so, the postulation of the faculty of reasoning is made central and intrinsic for cogency in practice. This move vindicating the locus of reason in mathematical practice seems to imply naturalizing deductive capacities. Moreover, this line of argument could allow a consistent articulation of a view rendering logic in some sense a posteriori and in some other sense a priori.

It follows that the normative *motto* that reasoning is "ruled by norms" is induced from rhetorical hyperbole or from trespassing the boundary between epistemics and pragmatics. In its present characterization, normativity appears as extrinsic or derivative to the cogency proper study. This is not to deny that the pragmatic dimension of reasoning can be "normalized" as a code of good practices giving rise to a kind of *clinic* normativity in the way in which sound reasoning "should" be presented in social and rhetorical argumentative practices. Barwise and Etchemendy (1999/2003: 48-49) identify this rhetorical element in the formal presentation of a proof:

> "The difference between an informal proof and a formal proof is not one of rigor, but of style. An informal proof of the sort used by mathematicians is every bit as rigorous as a formal proof. But it is stated in English and it is usually free-wheeling, leaving out the more obvious steps. A formal proof, by contrast, employs a fixed stock of rules and a highly stylized method of presentation".

These authors acknowledge that, "for purposes of proving things for ourselves, or communicating proofs to others, informal methods are usually preferable", in agreement with Myhill (1951: 70) and others. *Prima facie* this viewpoint is consistent with Feferman's idea that good thinking is to be found in pre-formal settings. Notice Barwise and Etchemendy do not go as far as Feferman in claiming that the ability for good thinking is a prerequisite for the study of formal logic. On the other hand, they imply that rigor is a feature of both informal and formal proofs. Of course, they do not explain what they mean by 'rigor' in the present context but it is safe to assume that they have in mind precision of expression, surely a fundamental trait in any proof. Moreover, they do not explain what they mean by 'style'. It seems clear that 'style' in their sense cannot be merely reduced to some sort of superficial difference in the, let us say, rhetorical

presentation of a given proof, since—as they claimed—in an informal proof sometimes *obvious steps are left out*.

Barwise and Etchemendy open up a Pandora box by holding the rather controversial viewpoint that enthymematic informal proofs are as rigorous as non-enthymematic formal proofs. At face value these authors, perhaps inadvertently, are suggesting the existence of two kinds of proofs, a viewpoint that appears to contradict the common place conception that an enthymematic *proof* contains a flaw in reasoning and hence it is no proof, in the same way that a silver bird is not a bird. They insist that they do not want to give the impression that formal proofs are somehow better than informal proofs. However, they seem oblivious about the historically well identified fallacy of smuggling premises. If Suppes is correct, the difference between a formal and an informal proof cannot be one of mere style since Suppes most probably would take enthymematic reasoning—reasoning in which some steps are left out—to be incorrect from the viewpoint of the model.

It is worth recalling that one historically recognized advantage of the use of formal logic is the resort to disinterpretation and re-interpretation of the language by substituting a given class of content terms by suitable variables. This method has been historically used to detect invalidity. On the other hand—as suggested above—it is doubtful that a student smuggling a premise or not understanding a given step of deduction can really come to see the problem by appealing to a formal counterinterpretation, much less to a rule of derivation. Notice that a formal counterinterpretation will show that certain piece of reasoning has "another" argument in the same logical form with true premises and false conclusion. This is not telling the thinker what actually went wrong in the reasoning. Similarly, there is no way to show incorrectness of a step of reasoning by pointing at a rule of inference of a different form. If this is right then a generalization of the Suppes' viewpoint would amount to put the chart before the horse when relying on a formal definition of proof to shed light on what is wrong or controversial in an informal proof. This analysis answers Shapiro's dilemma negatively. As a rule, we do not correct mathematical practice by appealing to a descriptive formal model of logic. Thus, the view of logic as model leaves us even, after all, with natural sciences common sense idea that "reality" is not to be blame by disconfirmation of the model.

Final remarks: models of physics and models of logic

I have suggested that with a few qualifications the analogy between models of physics and models of logic turns out to be more positive than negative. In particular, once we understand that it

is not a model *per se*, but a certain use of a model which may carry normative force, for example in an teaching or instruction context, then the view of logic as model is no more normative than any other particular scientific activity in which models are essential for teaching and research. This is pragmatics.

There is a natural rhetoric emphasis when saying that a particular subatomic setting *ought* to follow the planetary model or that a particular implication relation *ought* to follow semantic rules. These modal usages seem to reveal epistemic uncertainty on the side of the speaker rather than being genuinely normative. Likewise it is natural to adopt the normative mode when instructing a person in facing facts as experienced in a given practice and in experiencing the use of models to explain and predict those facts. What is meant is simply that under the normal course of things, if the model is right and the performance of the user of the model is right things turn out to be as they "should". The normative mode may be necessary sometimes, but it is certainly not sufficient for prompting understanding and thinking. On the other hand, we should not miss that we are often normative when correcting a bad performance or when guiding what we want to be a correct performance in the *very same muddle* in which a mistake can take place. This is the point where being a mathematician or being a logician does not make any difference. Instruction towards developing a certain goal or competence requires experience which cannot be obtained without practice and performance. When grasping the phenomena under consideration we realized that normativity takes place in pragmatically articulating means towards ends. Laws of nature and laws of logic exhibit a different kind of necessity none of which is normative necessity. Instruction towards knowledge of each can be regimented by norms and rules. But this is the realm of pragmatics, neither ontics nor epistemics. Shapiro's dilemma is instructive in many ways. For one lesson I take we cannot leave objectivity to be established by dogma or self-deception but only by hard work.

Acknowledgments: This paper is framed under research projects FFI2008-00085, y FFI2009-08828 of Ministerio de Ciencia e Innovación. I thank Concha Martínez and José Ferreirós for useful discussion.

REFERENCES

Barwise, J. and Etchemendy, J. 1999/2003, *Language, Proof and Logic,* Stanford, CSLI publications.

Boole, G. 1854/1958, *The Laws of Thought,* New York, Dover.

Corcoran, J. 1973, Gaps between logical theory and mathematical practice, *The Methodological Unity of Science,* Mario Bunge editor. Dordrecht-Holland, Reidel, 23-50.

Corcoran, J. 1992, Unprovability and undefinability. *Travaux de Logique,* D. Miéville editor, University of Neuchâtel, 37- 65.

Falguera, J. 1994, 'Unidad de noción bajo los usos del término 'modelo' en las ciencias matemáticas y factuales', *Contextos* XII/23-23, 221-244.

Feferman, S. 1964, *The Number Systems*. *Foundations of Algebra and Analysis,* Addison-Wesley.

Martínez, Concha 2001, The View of Logic as Model: Comments on Stewart Shapiro's paper. *Ágora* 20/1, 175-179.

Myhill, J. 1951, On the Ontological Significance of the Löwenheim-Skolem Theorem. In *Academic freedom, logic and religion* (M. White editor), 57-70.

Sagüillo, J. 2006, Argumentative practice: cogency and normativit,. *Actas del VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España,* Granada, Universidad de Granada.

Shapiro, S. 2001, Modeling and Normativity: How much revisionism can we tolerate?, *Ágora* 20/1, 159-173.

Suppes, P. 1957/99, *Introduction to Logic,* New York, Dover.

Von Wright, G. H. 1963/1971, *Norm and Action,* London, Routledge & Kegan Paul, New York, The Humanities Press.

II SEMINARIO ALFREDO DEAÑO SOBRE RAZONAMIENTO ORDINARIO

ORGANIZAN: A. SOBRINO & M. PEREIRA-FARIÑA

El razonamiento ordinario es aquel que usamos normalmente los humanos para argüir, debatir o resolver problemas cotidianos verbalizados en lenguaje común. Se distingue así del razonamiento formal o matemático, utilizado para garantizar el tránsito de la verdad de las premisas a la conclusión, asumiendo que esa verdad puede ser definida de modo preciso y definitivo.

Alfredo Deaño fue quizás el primer lógico español en divulgar la lógica del lenguaje ordinario en su conocido manual, *Introducción a la lógica formal* (1974). El texto de Deaño usa con profusión y claridad ejemplos del lenguaje común para ilustrar nociones y conceptos formales, tanto clásicos como alternativos, con el objetivo de mostrar que si la lógica quiere ser un asunto humano, tiene que acercarse al lenguaje cotidiano. Deaño apunta ya en su obra algunas de las limitaciones de los formalismos clásicos cuando se aplican a la semántica o a la forma del lenguaje común, haciendo ver, ya en aquel tiempo, la conveniencia de desarrollar otras lógicas que pusiesen cierto orden en el prolijo, complejo y resbaladizo mundo de la argumentación ordinaria.

Con este Simposio pretendemos recuperar y actualizar las preocupaciones de Alfredo Deaño convidando a investigadores de diferentes áreas de conocimiento, como la Filosofía, la Inteligencia Computacional, la Lingüística o la Psicología, a debatir y reflexionar acerca de las limitaciones, logros y retos por venir en el campo del tratamiento del razonamiento común, área interdisciplinar en la que confluyen variados saberes y diversas perspectivas.

Por acabar, decir que este simposio quiere dar continuidad a la primera edición celebrada en 2012 en el Centro Cultural ‘Muralla Romana’ de Gijón, en el que participaron investigadores relevantes del campo de la Filosofía, como Javier Muguerza, y de la Inteligencia Computacional, como E. Trillas y R. López de Mántaras.

ALGUNOS USOS DE LOS CONDICIONALES Y SU PROCESAMIENTO AUTOMÁTICO

SANTIAGO FERNÁNDEZ LANZA
USC

ABSTRACT: In this paper I describe a range of uses that human beings give to the expressions of type " if… then…" when everyday language is used. On many occasions, this type of statements represent conditionals, whose study has been fruitful in various fields, and now also it is interesting to Computer Science since " if... then..." (" if… then…") expressions constitute one of the main flow control statements in the context of programming languages. The analysis of conditionals, both aimed to identification automatically and its semantic interpretation by a computer, is interesting for what is now known as Computing with Words (Zadeh 1999). This is a multidisciplinary and complex work and, therefore, this paper is a half way among Linguistics, Philosophy of Language, Logic and Computer Science. It may be interesting in itself in any of these four areas, provided that considers relevant observation of the use of Spanish speakers make expressions "If the type…then…".

Keywords: Conditionals, Identification of conditionals, Evaluation of conditionals, Computing with Words.

1 Introducción

Desde los estudios de Lógica en la Grecia Antigua hasta nuestros días se han realizado diversas propuestas de caracterización, clasificación y evaluación de condicionales. Con relativa frecuencia, estas propuestas han ido encaminadas hacia su evaluación, es decir, hacia el establecimiento de las condiciones bajo las que podemos catalogar a un enunciado condicional como verdadero, falso, etc.

En (Fernández-Lanza 2008) se trata de mostrar en qué medida la evaluación de los condicionales propuesta por la Lógica Formal responde a las intuiciones y modos de operar de los humanos a la hora de evaluarlos como verdaderos o falsos. La citada propuesta hace depender el valor de verdad del condicional completo de los valores de verdad de antecedente y consecuente. Mediante un pequeño experimento empírico, también descrito en el citado artículo, se pretende mostrar que tal dependencia no suele tener lugar en la evaluación de enunciados condicionales que los humanos realizamos. El experimento parece indicar que los humanos pueden evaluar condicionales como verdaderos o falsos a partir de cualquier combinación de valores de verdad de antecedente y consecuente excepto en el caso de que el antecedente sea verdadero y el consecuente sea falso, puesto que, parecemos estar de acuerdo en que, para esta combinación, el condicional es falso.

2 Identificación de condicionales

Con el surgimiento de las Tecnologías de la Información y las Comunicaciones han aparecido contextos donde resulta interesante el estudio de una cuestión previa a la evaluación de condicionales pero no por ello menos importante: su identificación. En el ámbito del Procesamiento del Lenguaje Natural se han establecido propuestas de tratamiento de condicionales basadas fundamentalmente en la localización de expresiones como "si", "cuando", "entonces", etc. Es decir, *grosso modo* se considera a un enunciado condicional como aquel que contiene una o varias de las citadas expresiones dispuestas o intercaladas en el texto de determinada manera. Los investigadores de este ámbito son conscientes de las limitaciones que tienen este tipo de propuestas y los errores que frecuentemente acarrean, por este motivo, se añaden a estos procedimientos una serie de estrategias correctoras que, en ocasiones, proporcionan resultados relativamente satisfactorios. En este artículo se entenderá la identificación de un condicional como el proceso mediante el cual los humanos reconocemos que determinado enunciado es un condicional o no lo es. Para un lógico este proceso es relevante a la hora de formalizar un enunciado. Para los investigadores del campo del Procesamiento del Lenguaje Natural, este proceso se simula mediante el uso de estrategias de localización de expresiones como la que se acaba de describir anteriormente de forma excesivamente breve y esquemática. Sin embargo, este proceso no resulta adecuado principalmente en dos ocasiones:

- La primera de ellas tiene lugar cuando los condicionales están expresados sin hacer uso de los términos lingüísticos que habitualmente se utilizan para ellos. Por ejemplo, un enunciado como "Termina el proyecto y te subo el sueldo" puede formalizarse como la conjunción de los dos enunciados simples "Termina el proyecto" y "Te subo el sueldo", pero la realización de un análisis más fino revela que el enunciado responde también al condicional "Si terminas el proyecto entonces te subo el sueldo". En este caso, una máquina que utilice el procedimiento de simple comparación de cadenas de texto dejaría de identificar un condicional debido a que no ha localizado en el enunciado expresiones como "si", "entonces", etc.
- En segundo lugar, la estrategia de localización de expresiones no es adecuada cuando se utilizan las expresiones condicionales para dar cuenta de enunciados que podrían no ser interpretados como condicionales. El procesamiento automático de este tipo de enunciados identificaría un condicional que podría no ser considerado como tal.

2.1. Distintas interpretaciones de un enunciado con aparente formato condicional

Respecto a este segundo punto, en (Fernández-Lanza 2010) se describen dos experimentos empíricos, el primero de los cuales trata de mostrar que ciertos enunciados con aparente formato

condicional pueden tener varias interpretaciones para diversos hablantes de una lengua. El segundo de ellos pretende poner de manifiesto si esta diversidad de interpretaciones puede influir en los sujetos a la hora de establecer el valor veritativo de los enunciados originales. Los enunciados utilizados en esos experimentos fueron:

1. *Si George Bush es pacifista entonces yo soy un marciano*
2. *Si tienes hambre, hay galletas en la alacena*
3. *Si hay algo que no comprendo es que te guste la carne cruda*
4. *Si te has matriculado entonces podrás ir al examen*
5. *Si George Bush es estadounidense entonces es norteamericano*

Ante las distintas alternativas, parece que las interpretaciones más populares para los ejemplos 1, 2 y 3 son aquellas que no identifican al enunciado con un condicional, mientras que para los ejemplos 4 y 5, sus respectivas interpretaciones condicionales son las más seleccionadas. Si el experimento descrito se pudiese extrapolar a toda una comunidad de hablantes de una lengua, se podría indicar que la mayoría de los hablantes de español no entienden el enunciado "Si George Bush es pacifista entonces yo soy un marciano" como un condicional sino como "George Bush no es pacifista" o "El hablante afirma que George Bush no es pacifista"; el enunciado "Si tienes hambre, hay galletas en la alacena" como "El hablante está ofreciendo galletas a el oyente" o bien "Hay galletas en la alacena" y el enunciado "Si hay algo que no comprendo es que te guste la carne cruda" como "El hablante no comprende que al oyente le guste la carne cruda" o bien "El hablante afirma que no comprende que al oyente le guste la carne cruda". Estas interpretaciones podrían tener una ventaja a la hora de evaluar tales enunciados ya que las alternativas interpretadas son, en principio, más sencillas de evaluar que los enunciados condicionales originales. Sin embargo, el segundo de los estudios empíricos no parece revelar ningún indicio claro de que las distintas interpretaciones que se pueden dar de los cinco enunciados analizados, puede afectar a la evaluación de éstos como verdaderos, falsos, etc. por parte de los sujetos.

2.2. Disparidad de usos de las expresiones condicionales

El lenguaje natural es lo suficientemente complejo como para permitirnos la utilización de expresiones condicionales con distintos objetivos, cada uno de los cuales puede hacerse saber al interlocutor de una forma más o menos explícita mediante el uso de determinadas palabras clave dentro del propio enunciado. Comúnmente, los humanos utilizamos los condicionales cuando razonamos. En este contexto se pueden utilizar explícitamente expresiones indicadoras del tipo de argumento en el que se podría incardinar el condicional que proferimos. De esta forma podemos decir:

1. Si estoy de pie y tengo las gafas puestas entonces necesariamente tengo las gafas puestas.
2. Si es español entonces probablemente es de raza blanca.
3. Si estás comiendo chiles entonces conjeturo que se te ha curado la úlcera de estómago.

No lo hacemos con mucha frecuencia pero cuando lo hacemos al oyente le queda más claro lo que le cabe esperar si el antecedente se satisface.

Otro contexto, relacionado con los anteriores, en el que utilizamos los condicionales es para hacer predicciones del estilo "Si mañana llueve entonces Alonso ganará la carrera". Podríamos expresar este tipo de condicionales de la siguiente forma:

4. Si mañana llueve entonces predigo que Alonso ganará la carrera.

También utilizamos condicionales para enunciar reglas o leyes ya sean leyes de la naturaleza, leyes jurídicas etc. como por ejemplo "Si se calienta agua a más de 100º entonces hervirá" o "Si estás matriculado entonces puedes examinarte". Algo que podemos expresar de la siguiente forma:

5. (a) Si se calienta agua a más de 100º entonces (naturalmente) hervirá.
   (b) Si estás matriculado entonces (legalmente) puedes examinarte.

Cuando hacemos instancias de reglas generales con la forma "Todos los A son B" como por ejemplo "Todos los españoles son europeos" solemos hacerlo de forma condicional diciendo "Si soy español entonces soy europeo". Por eso su instancia podría expresarse como:

6. Si soy español entonces (como todos los españoles son europeos) soy europeo.

Otro uso de enunciados con aparente formato condicional es el que tiene lugar cuando los utilizamos como recurso para expresar la negación de su antecedente o, dicho de otra forma para dar a entender la negación de lo que dice el antecedente (esto es, ironizar), como sucede cuando para indicar que el oyente no es el presidente del gobierno se utilizan oraciones como "Si eres el presidente del gobierno entonces yo soy el rey de España". Estos condicionales se suelen denominar condicionales contra-fácticos indicativos. Quizás para formalizar enunciados como este, lo más aconsejable sería hacerlo como la negación de su antecedente y no como un condicional. En cierto modo en estas expresiones aparentemente condicionales se podrían incluir etiquetas como las señaladas en el siguiente ejemplo:

7. Si eres el presidente del gobierno entonces irónicamente yo soy el rey de España.

También utilizamos condicionales para enfatizar lo expresado por el consecuente. Como cuando un individuo, para poner énfasis en el hecho de que le molesta que lo interrumpan cuando habla, lo expresa diciendo "Si algo me molesta es que me interrumpan cuando hablo". Son los llamados condicionales con antecedente trivialmente verdadero. Una peculiaridad de este tipo de con-

dicionales es que resulta necesario parafrasearlos para que posean un formato del tipo "Si … entonces …" de la siguiente manera "Si algo me molesta entonces me molesta que me interrumpan cuando hablo". La versión incluyendo palabras clave podría ser:

8. Si algo me molesta entonces (enfatizo que) me molesta que me interrumpan cuando hablo.

Otro uso de los condicionales es el caso de las amenazas condicionadas, como por ejemplo en "Si cuentas esto entonces te mataré". Las etiquetas incorporadas a este ejemplo hacen explícito el acto de habla de la siguiente forma:

9. Si cuentas esto entonces amenazo con que te mataré.

También utilizamos los condicionales para prometer cosas como en "Si terminas el proyecto entonces te subiré el sueldo". La palabra clave es también la que hace explícito el acto de habla:

10. Si terminas el proyecto entonces prometo que te subiré el sueldo.

Solemos utilizar expresiones aparentemente condicionales para ofrecer cosas. Los *biscuit conditionals* de Austin ("Si tienes hambre, hay galletas en la alacena") son un ejemplo de ello [véase (Austin 1956, 109-132)] y su versión explícita podría expresarse de la siguiente forma:

11. Si tienes hambre entonces te ofrezco las galletas que hay en la alacena.

Cuando damos consejos, ordenes o peticiones también podemos hacerlo mediante el uso de condicionales como en "Si vas a salir entonces lleva paraguas", "Si vas a llegar tarde entonces no vuelvas a casa" o "Si vas a la cafetería entonces tráeme tabaco" respectivamente. En estos casos, cuando tenemos poca información contextual puede ser difícil distinguir lo que es un consejo de una orden o una petición. Este es el motivo por el que, con cierta frecuencia hacemos explícito el acto de habla en estos casos:

12. (a) Si vas a salir entonces te aconsejo que lleves paraguas.
    (b)Si vas a llegar tarde entonces te ordeno que no vuelvas a casa.
    (c) Si vas a la cafetería entonces te pido que me traigas tabaco.

La identificación de cadenas de texto en un pasaje lingüístico es un problema relativamente simple desde el punto de vista computacional dados los actuales avances de los lenguajes de programación. La detección de las expresiones clave señaladas en este apartado puede llevarse a cabo sin excesiva complicación. El problema es que no siempre expresamos lo mismo con las mismas palabras, por este motivo un diccionario electrónico de sinónimos como el presentado en (Fernández-Lanza 2002) puede ser de gran ayuda en esta ocasión. Con el uso de tal diccionario se pueden multiplicar las expresiones clave teniendo en cuenta el grado de sinonimia con respecto a la expresión clave estándar.

2.3. Los tiempos verbales

La mera identificación de partículas como "si", "entonces", "cuando", etc. no basta para identificar un enunciado como condicional por los motivos relatados al principio del aparatado 2. Son necesarias, por tanto, estrategias computacionales que ayuden en tal identificación. Parecen existir algunos criterios que permiten obtener más información sobre los usos que hacemos de los condicionales. Uno de ellos, es la localización de determinadas expresiones clave como "necesariamente", "probablemente", "te aconsejo que", etc. dentro del enunciado, tal y como se ha indicado en el apartado 2.2. Este criterio no siempre es efectivo ya que no siempre hacemos explícito el uso del condicional que estamos utilizando. Un criterio complementario que puede ser de utilidad en caso de fallo del criterio de las expresiones clave podría ser atender a los tiempos verbales de antecedente y consecuente, ya que no parecen ser válidas todas las combinaciones en todos los usos.

Una regla que parece operar de forma general en los condicionales es aquella que indica que no es usual utilizar tiempos futuros en el antecedente de un condicional sino perífrasis verbales de futuro, (decimos "Si vas a salir entonces deberías llevar paraguas" y no "Si saldrás entonces debes llevar paraguas")[1]. En (Fernández-Lanza 2012) se realiza un estudio de los 12 usos de condicionales ejemplificados en el apartado 2.2, en el que se concluye con el establecimiento de las siguientes reglas generales respecto a la tolerancia a cambios de tiempo verbal entre antecedente y consecuente:

- Los usos con el adverbio "necesariamente", las instancias de reglas generales y los usos para enfatizar suelen precisar de coincidencia de tiempo verbal entre el antecedente y el consecuente.
- Los consejos, órdenes o peticiones suelen admitir cualquier tiempo en el antecedente y únicamente presente de imperativo o subjuntivo en el consecuente.
- Las predicciones, las amenazas y las promesas no suelen admitir tiempos pasados en el consecuente.
- Las leyes suelen expresarse con tiempos en el antecedente menores o iguales a los del consecuente.

[1] No debe confundirse el uso de la partícula "si" como condicional y el uso de la partícula "si" como conjunción interrogativa o dubitativa que introduce oraciones sustantivas o interrogativas indirectas como en "si vendrá, no está claro" o "no está claro si vendrá" [para una aclaración sobre esta cuestión véanse (R.A.E. 1973), pp. 520-522 y (Gómez-Torrego 1997, 322-327)]

- Para los usos con el adverbio "probablemente", las conjeturas, cuando los utilizamos para ironizar o cuando ofrecemos (*biscuit*) no tenemos restricciones con respecto a los tiempos verbales empleados.

Ordenando los ejemplos de menor a mayor grado de tolerancia la tabla 1 resume el citado estudio[2]. Si computacionalmente somos capaces de detectar el tiempo verbal del antecedente y el consecuente de un condicional, la tabla 1 puede, para algunas combinaciones, ayudar a detectar el tipo de uso que estamos haciendo de éste cuando no se utilizan expresiones que hagan explícito dicho uso.

Tabla 1: Tolerancia a cambios de tiempo verbal en condicionales.

Ant	**Cons**	**Nec**	**Reg**	**Enf**	**Csj**	**Ame**	**Prm**	**Prd**	**Ley**	**Prb**	**Cnj**	**Iro**	**Ofr**
PAS	PAS	1	1	1	0	0	0	0	1	1	1	1	1
PAS	PRES	0	0	0	1	1	1	1	1	1	1	1	1
PAS	FUT	0	0	0	0	1	1	1	1	1	1	1	1
PRES	PAS	0	0	0	0	0	0	0	0	1	1	1	1
PRES	PRES	1	1	1	1	1	1	1	1	1	1	1	1
PRES	FUT	0	0	0	0	1	1	1	1	1	1	1	1
FUT	PAS	0	0	0	0	0	0	0	0	1	1	1	1
FUT	PRES	0	0	0	1	1	1	1	0	1	1	1	1
FUT	FUT 1	1	1	0	1	1	1	1	1	1	1	1	1

3 Conclusiones

Uno de los principales objetivos de la *Computación con Palabras* es el de tratar de utilizar el lenguaje natural para la implementación de algoritmos y funciones lógico-computacionales. Hasta el momento estos algoritmos y funciones se expresan mediante el uso de lenguajes de programación. Una sentencia de control de flujo fundamental en tales lenguajes son los "if … then …" ("si … entonces…"), es decir, sentencias con formato condicional que permiten la ejecución de ciertas instrucciones en función del cumplimiento o no de determinadas condiciones. Conseguir que una máquina sea capaz de procesar sentencias con esa estructura a partir de lenguaje tal y como los humanos lo expresamos, debería requerir el estudio de procedimientos de identificación e interpretación de tales sentencias. Por este motivo, en el campo de la Computación resulta interesante el análisis de las expresiones condicionales y su procesamiento automático.

El presente trabajo ofrece un intento de análisis basado en el uso que los humanos le damos a los condicionales. Se han citado estudios empíricos con hablantes de español mediante los que se

[2] Leyenda: Ant=Antecedente, Cons=Consecuente, Nec=Necesariamente, Reg=Reglas Generales, Enf=Énfasis, Csj=Consejos, Ame=Amenazas, Prm=Promesas, Prd=Predicciones, Ley=Leyes, Prb=Probablemente, Cnj=Conjeturas, Iro=Ironizar, Ofr=Ofrecer

ha podido vislumbrar que la evaluación de los condicionales propuesta por la Lógica Formal responde parcialmente a las intuiciones y modos de operar de los humanos a la hora de evaluar los condicionales como verdaderos o falsos. Que la identificación de los mismos no es una tarea tan sencilla como detectar expresiones mediante la mera comparación de cadenas de texto. Que podemos tener sentencias que carecen de formato condicional ("si … entonces …") y sin embargo po drían ser identificadas como condicionales. Que cabe la posibilidad que una expresión con supuesto formato condicional pueda ser interpretada de varias maneras por un hablante y alguna de éstas pueda carecer de una estructura condicional. A este respecto, se han ejemplificado 12 tipos diferentes de condicionales que podrían ser identificables con relativa comodidad si el emisor fuese lo suficientemente caritativo como para intercalar determinadas expresiones clave. Pero, por economía del lenguaje, ésta no es una práctica habitual. Por este motivo se propone la estrategia de la observación de la tolerancia a cambios de tiempo verbal entre antecedente y consecuente, que aunque no es completamente efectiva, podría proporcionar ayuda en la identificación.

Agradecimientos

El presente trabajo ha sido financiado por los proyectos INOU11A-04 de la Diputación Provincial de Ourense y FFI2009-08828 del Ministerio de Ciencia e Innovación.

REFERENCIAS

Austin, J. L. 1956. "Ifs and Cans." *Proceedings of the British Academy* 42:109–132.

Fernández-Lanza, S. 2002. "Una contribución al procesamiento automático de la sinonimia utilizando Prolog." *Revista de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)* 28:109–110. .

Fernández-Lanza, S. 2008. "Una propuesta metodológica para la evaluación de condicionales." *Revista de Filosofía* 33(2):67–86. . 2010. "Dos experimentos sobre el uso de condicionales." *Agora* 29(2):55–78. . 2012.

Fernández-Lanza, S. "Un análisis para diversos usos de los condicionales" en *Ensayos sobre Lógica, Lenguaje, Mente y Ciencia*, editado por L. Fernández-Moreno. (En prensa).

Gómez-Torrego, L. 1997. *Gramática didáctica del español*. Madrid: Ediciones SM.

R.A.E. 1973. Esbozo de una nueva gramática de la lengua española. Madrid: Espasa-Calpe.

Zadeh, L. A. 1999. "From Computing with Numbers to Computing with Words: From Manipulation of Measurements to Manipulation of Perceptions." *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 45(1):105–119.

INTERPRETATIONS AS COGNITIVE EVENT TYPE

MARÍA G. NAVARRO
CSIC

ABSTRACT: In this paper it is argued that abductive reasoning is part of a cognitive process that involves interpreting something as being something. In order to do so, a definition of proposition and interpretation as type of cognitive events is proposed. Taking this into account, it is established that the inferential pattern corresponding to this type of cognitive event is abductive as, in some cases, the activity of interpreting is a process of searching for plausible assumptions to form consistent explanatory hypotheses.

Key words: Proposition; Scott Soames; Cognitive States; Abductive Reasoning.

We may say that propositions are crucial to any cognitive theory. Propositions have a life beyond language since they are partly a consequence or result of the content of cognitive states. Propositions thus have the peculiarity of fulfilling different functions simultaneously: the linguistic function, the perceptual function and the function of representing cognitive content. It is due to their possibility of simultaneously fulfilling these three functions that Scott Soames (2010) understands propositions as minimal units of their representational type.

Soames (2010, 2012 forthcoming) has proposed a definition of the proposition that extends it into a type of cognitive event. Propositions are often parts of large cognitive structures that exhibit some of the holism and counterfactual interconnection exhibited by perceptual states. In *What is Meaning?* Soames puts forward a theory about the proposition as a type of cognitive event based on two fundamental ideas: (i) that the perceptual and cognitive activity of agents is the conceptual basis of all representation and (ii) that propositions are representational in virtue of the relations they bear to this representational activity. In this article we shall assume Soames's notion of the proposition in order to define the process whereby we interpret something as something.

A proposition is a type of entity that is not identical with the sentence that expresses it, since the same proposition may be expressed by different sentences. Propositions have to do with the contents of perceptual and cognitive states. However, the difficulty reappears at this point because, as in the case of sentences, these contents cannot be baldly identified with propositions. With propositions we represent the world to ourselves as being a certain way, but it is precisely by using propositions that we manage to represent it otherwise. If propositions, from the semantic standpoint, establish the possibility of truth conditions (if they are a kind of truth and/or falsity bearers), this is be-

cause they are not actually identical with the semantic contents of sentences. Despite these difficulties that arise when we attempt to determine what kind of entity a proposition is, the traditional approach continues to insist that they are partly sentences and partly the contents of our perceptual and cognitive states and consequently bearers of truth and/or falsity.

The conception of the proposition as a type of cognitive event presents this as something more than the mere effect of cases in which we represent things as being in this way or that. Soames sets out this idea in the following hypothetical case: the proposition that o is red is the event type in which an agent predicates redness of o, it represents o as being red because all conceivable instances of it are events in which an agent does so.

This leads Soames to claim that the minimal form of knowledge with propositions stems from their capacity to represent the world cognitively or perceptually by predicating properties of objects and thus generating proofs of types of events corresponding to those predications. To obtain a more solid form of knowledge, one must be able to construct propositions that are (and transform them into) an object of predication.

Under his conception of the proposition Soames understands the action of predicating as a type of cognitive event, and this implies that such an action is analogous to the intensional transitive verb "to seek". When a proposition is the event type of predicating a property of an object o, o may be a constituent of the proposition —in the sense that the proposition is defined in terms of o— without o´s existence being necessary for the existence of the proposition.

Both propositions and sentences are types of complex events that involve the performance of certain basic acts. What guarantees the existence of propositions? The answer seems as simple as it is radically original: whatever guarantees the existence of events in which those acts are performed. This means that many propositions have not and never will be considered.

Propositions are assumed and used in many types of events where agents perform certain operations of cognitive representation. Traditionally, it is understood that an interpretation is the product or result of the act of interpreting. The act of interpreting can be distinguished from its products, that is, from the interpretations. Interpretations, understood as a product, are expressed in propositions. Following Soames, we may establish a parallel between what is traditionally understood by *proposition* and what is commonly understood by *interpretation*. It is generally accepted that proposition and interpretation are entities that fulfil A1-A3.

A1. Some things are interpreted. These attitudes relate those who interpreted something to that which they interpreted.

A2. The things interpreted are bearers of truth and falsity.

A3. Interpretations —the things satisfying A1 and A2 are expressed by sentences. Interpretations expressed by S can be designated by the expressions [the proposition that S] , [the statement/claim/assertion/belief that S] , or simply [that S].

In this article we propose a definition of the act of interpreting according to which this activity produces a specific type of events called "interpretations" that preform the cognitive act of interpreting. In effect this implies that many interpretations have not yet been expressed. Just as a proposition is an entity that is not identical with the sentence that expresses it – the same proposition may be expressed by different sentences – we maintain that the act of interpreting is a type of cognitive event that cannot be identified with its products alone: interpretations. There are propositions that could be presented finally as the product of this type of cognitive event (its resulting interpretation). In the same way that propositions are not identical with the sentences with which they are expressed (there are different sentences or expressions that express the same belief), it may be said that interpretations are not to be identified with sentences.

Semantics needs propositions to specify semantic content: that which is asserted and believed by speakers. The same happens with interpretations from the semantic standpoint. Traditionally, interpretations are a specific type of entity that represents the world as being a certain way, which is why it has often been said, and it is even believed, that they have truth conditions.

When it is not associated with philosophical currents such as hermeneutics or even with a continental vision of philosophical activity (in contrast to analytic), the problem of interpretation is usually approached either from a semantic or from a pragmatic position (Dascal, 1995; Gabbay and Woods, 2005; González Navarro, 2009, 2011, 2012b). In this sense, a paradigmatic example is that of Bunge (1974), who distinguishes interpretation as a semantic concept from interpretation understood in terms of its reference to social facts. Although the arguments Bunge uses to support the separation between semantic interpretation and interpretation as a specific problem of the social sciences do not convince us, we share the idea of semantic interpretation referring to signs and constructs. According to this usage, the meaning of a construct should not be confused with the mental process of understanding, that is, an interpretation should not be confused with the mental activity of interpreting.

The products of the activity of interpreting do not get their representational character from the sentences that express them (King, 2007), but rather from the fact that the interpretations (as we maintain here) are the content of intentional cognitive states.

What fact makes it legitimate to attempt to go beyond the semantically and/or pragmatically accepted meaning of the interpreting activity? Why is the interpreting activity defined here as a type

of cognitive event? We usually choose an interpretation expressed by a proposition *p* when it is part of the content of a perceptual or cognitive state we wish to characterise. That is the reason Soames asserts that we use sentences to express complex cognitive contents that represent other cognitive or perceptual contents as satisfying various conditions.

As we are seeing, this analysis of the interpreting activity is not confined to either semantic or pragmatic accepted usage. We propose here to analyse the activity of interpreting as a type of cognitive event that takes on the form of abductive reasoning. However, to define the interpreting activity as a type of cognitive event expressed by abductive reasoning procedures clearly demands that the activity be placed not only in the representational but also in the inferential tradition.

According to Brandom (1994), enlightenment epistemology accommodated two different conceptions of the origin, order and connection of ideas. On the one hand, there was the notion of representation espoused by Descartes, who divided the world into those who by nature are subjects of knowledge and make representations (*res cogitans*), and that which is the object of knowledge and constitutes a representation (*res extensa*). In doing this, Descartes made the act of possessing contents of representation an explanatory instance, but at the same time made it inexplicable.

By contrast with Descartes, both Leibniz and Spinoza wondered what it is that indicates that one thing represents another, but they did this in terms of the inferential meaning of what represents, which should, in any case, have been explained by inferential relations. As Brandom reminds us, the British empiricists also claimed that inferential relations should be extracted from the contents of representations, since a subject's mere representation of something does not inform other subjects as to whether his representation is indeed a representation *of* something. It was then that inferential correctness was interpreted on the basis of the correctness and intelligibility of the representation or representations. Since then, to a great extent, philosophers have attempted to define representational properties in terms of inferential ones.

In the history of the philosophy of logic, the definition and analysis of syntactic and semantic problems have constantly been related to representational and to inferential properties. For example, it is usually said that syntactic problems are related to valid inference, to the concepts of demonstration and derivation and to the axiomatic method; while semantic problems relate to analysis of the concepts of meaning, truth, the distinction between intension and extension, and the paradoxes and antinomies. The problem of the logic of interpretation does not occupy such a salient position in the philosophy of language as that of representation or inference. Nevertheless, the logic of interpretation is related with the two latter concepts. Indeed, the problem of interpretation is the

means of resolving some paradoxes associated with cognoscitive activity dealt with by the representationist and inferentialist tradition.

The Aristotelian and mediaeval tradition, having a keen interest in the distinction between syntactic and semantic problems, already differentiated the *logica utens* from the *logica docens* (Beuchot, 2000). Hermeneutics would consequently conform to logic with regard to *corpus*, which could be, at the same time, the art and arrangement of procedures capable of being applied in all reasoning. This is why, in ancient times, it was accepted that hermeneutics was concerned with the application of methodological rules and principles or the *logica utens*, but also with the analysis of its content with regard to interpretation theory or *logica docens*. Centuries later, in spite of the theoretical separation accentuated by the analytic philosophical mindset versus the continental tradition, the epistemologists and philosophers of language assume that neither mere representations nor our conceptions about inferential validity can consistently explain the phenomena of understanding and reasoning. This situation has led some authors to criticise the so-called Standard Analytic Epistemology (Bishop and Trout, 2005).

For every interpretative process there is a piece of abductive reasoning that clarifies the logical schema of that process. To understand when the use of abductive reasoning comes into the cognitive activity of interpreting, we must ask ourselves: what does *interpret* mean? or: what do we do when we interpret? Interpretation is a complex phenomenon made up of two wholly integrated stages, although they are not easily distinguishable. In the first, after an act of comprehension, a proposition is formulated in which something is taken as something. Subsequently, the correlation and appropriateness of what is asserted is examined from a semantic point of view in relation to a set of propositions that in some cases may form part of a theory

To postulate the existence of these two stages in the phenomenon of the act of comprehension, that is, in the phenomenon of interpretation proper, entails that if we accept that interpretation implies taking something as something or in so much as something, then no proposition could include, in itself, the conditions under which one accepts its comprehensibility, since that act of comprehension could not be identified, in fact, with any form of totality. This is what the tradition of philosophical hermeneutics referred to with the idea that historical consciousness goes beyond its own knowledge of itself.

The acquisition of a language and the historicity of our comprehension preform our cognition through time individually and collectively. Here we have maintained that the ampliative effects observed in the inference to the best explanation are the result of applying a model of reasoning that is integrated into the act of interpreting. The inferential parameters that determine the logical rela-

tion between *explanandum, explans* and abductive explanation are inseparable from abductive competence understood in a dual sense: the theoretical and the manipulative. To live interpreting seems to be the phenomenon implicit in that model and pattern of reasoning, but what is certain is that only by immersing ourselves fully in it can we discover the complexity inherent in that fundamental mode of being in the world. That interpretation is defined here as a type of abductively formed cognitive event implies that this mode of being in the world must be investigated on the basis of its integration into the study about the so-called Ordinary Reasoning (Etherigton, 1988; Larrazabal and Pérez Miranda, 1992) or Commonsense Reasoning (Trillas, 2012), which is used by all living beings.

The definition of proposition proposed by Soames according to which it is a type of cognitive event has led us, firstly, to identify the activity of interpreting with that of producing such events from cognition. Secondly, we sustain the thesis that such an event constitutes a reasoning process modelled on an abductive pattern. Finally, as a consequence of this proposal, in presenting the said definition of interpretation we propose to introduce Soames's definition of proposition into a philosophical problem that is surely not broader but certainly different: that of philosophical hermeneutics where the idea of conversation provides and epistemic model (Adler, 2008). Soames himself noted how complex and significant is the existence of that border area (between the analytic and the continental) when he claimed that the structural relationships between the constituents of a proposition are given by the roles the constituents play in the sequence of cognitive operations performed by an agent who entertains it —roles like being predicated (of certain things), being targets (of certain predications), being applied (to certain arguments), being arguments (to which certain things are applied). These do not seem capable of being understood independently of the cognitive activity of interpreting. These pages offer some reasons to help understand why, as Mario Bunge (1974, 2009, p. 64) affirmed, interpretations do not arrive out of the blue; or remain, necessarily, once they have been put forward.

REFERENCES

Adler, J. 2008, 'Conversation is the Folks' Epistemology', *The Philosophical Forum* 39, 337-348.

Beuchot, M. 2000, *Tratado de hermenéutica analógica. Hacia un nuevo modelo de interpretación,* México, D.F., Ítaca.

Bishop, Michael A.; TROUT, J.D. 2005, *Epistemology and the Psychology of Human Judgment,* Oxford, Oxford University Press.

Brandom, R. 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment,* Cambridge, Harvard University Press.

Bunge, M. 1974, *Treatise on Basic Philosophy*. Vol. 2: *Semantics II: Interpretation and Truth*, Reidel Publishing Company.

Dascal, M. 1995, 'Epistemología, controversias y pragmática', *Isegoría. Revista de filosofía moral y política* 12, 8-43.

Etherigton, D. W. 1988, *Reasoning with Incomplete Information,* London, Pitman.

Gabbay, D. M. y Woods, J. 2005, *A Practical Logic of Cognitive Systems: The Research of Abduction,* New York, Elsevier Science Inc.

González Navarro, M. 2009, *Interpretar y argumentar*, Madrid/México, CSIC/Plaza & Valdés.

González Navarro, M. 2010, 'Intelligent Environments and the Challenge of Inferential Processes', *Tijdschrift voor Filosofie* 72(2), 309-326.

González Navarro, M. 2011, 'Hermenéutica', in Vega Reñón, L. and Olmos, P. (eds.), *Compendio de Lógica, Argumentación y Retórica*, Madrid, Editorial Trotta, 271-276.

González Navarro, M. 2012a, 'On Fuzziness and Ordinary Reasoning', in Rudolf Seising, R., Trillas, E., Moraga, C., Termini, S. (eds.), *On Fuzziness. A Homage to Lotfi A. Zadeh*, Springer, Berlin, New York, [et al.] (forthcoming)

González Navarro, M.; DE HARO GARCÍA, N. 2012b, 'Cognitive Abduction in the Study of Visual Culture', in Magnani, L. and P. Ling (eds.), *Philosophy and Cognitive Science*, Heidelberg/Berlin: Springer, Sapere vol. 2, 205-220

King, J. C. 2007, *The Nature and Structure of Content,* Oxford, Oxford University Press.

Larrazabal, J. and Pérez Miranda, L. A. 1992, 'Razonamiento Ordinario: Modelos mentales y modelos formales', *Contextos* X/19-20, 263-293.

Magnani, L. and Piazza, M. 2005, 'Morphodynamical Abduction. Causation by Attractors Dynamics of Explanatory hypotheses in Science', *Foundations of Science* 10, 107-132.

Magnani, L. 2009, 'Abducing chances in hybrid humans as decision makers', *Information Sciences* 179, 1628-1638.

Soames, S. 2010, *What is Meaning?* Princeton and Oxford, Princeton University Press.

Soames, S. 2011, "Propositions", in Gillian Russell and Delia Graff (eds.), *Routledge Companion to Philosophy of Language*, Routledge, Taylor & Francis Group.

Soames, S. 2012, "Propositions as Cognitive Event Type", in Jeff King, Scott Soames and Jeff Speaks (eds.), *New Thinking about Propositions*, forthcoming (under contract with Oxford University Press).

Trillas, E. 2012, 'A Model for '"Crisp Reasoning" with Fuzzy Sets', (manuscript)

UNA PROPUESTA DE SILOGISMO ENCADENADO APROXIMADO

M. PEREIRA-FARIÑA, A. SOBRINO
USC

ABSTRACT: Syllogism is a kind of deductive and demostrative reasoning that involves quantified statements, although limited to a few inference patterns and precise fragments of natural language. In this paper, we propose an alternative type of syllogism introducing the concept of analogy, evaluating the similarity of quantifiers through cardinality and developing a kind of reasoning characterized by plausibility.

KEY WORDS: Syllogism, Analogy, Similarity.

1 Introducción

La preocupación por el pensar ordenado y correcto ha sido y es una constante en la filosofía. La primera aproximación sistemática a una forma de razonamiento concreta fue la Silogística aristotélica (Aristóteles 1986), que data el siglo IV a.C. y que se caracteriza fundamentalmente por tratar con argumentaciones,

- compuestas por dos premisas y una conclusión, siendo todas ellas oraciones cuantificadas;
- que incluyen tres términos; el Término Mayor, que aparece en la Premisa Mayor y en el predicado de la conclusión; el Término Menor, que aparece en la Premisa Menor y en el sujeto de la conclusión y, por último, el Término Medio, que sólo aparece en las premisas;
- en las que sólo se contemplan los cuatro cuantificadores clásicos (*todos, ningún, algún, alguno… no*) habitualmente representados en el Cuadrado Lógico de la Oposición (Brown 1984).

También es relevante indicar que el silogismo es una forma de razonamiento deductivo; es decir, se determina que una conclusión se sigue necesariamente de ciertas premisas; tal y como lo define J. Corcoran en una de sus acepciones (Corcoran 2011). En la Tabla 1, mostramos un ejemplo típico de silogismo aristotélico.

Ejemplo 1. Razonamiento silogístico válido.

Todos los *aficionados a la música clásica* son melómanos

Algún *aficionado a la música clásica* es gallego

Algún gallego es melómano

Este silogismo constituye una argumentación en modo *Datisi* (Figura III) donde la conclusión respecto de las premisas está clara: cualquier subconjunto de *aficionados a la música clásica* también lo es del conjunto *melómanos*.

Esta sistematización del silogismo, sin embargo, excluye por completo a la vaguedad, lo que supone una limitación para su aplicación a fragmentos amplios del lenguaje natural. Por ejemplo, todos los cuantificadores vagos (como *mucho, bastante, el doble, pocos,…*), tan frecuentemente usados, quedan descartados de esta forma de razonamiento. También quedan fuera aquellas argumentaciones que contengan más de dos premisas y tres términos.

Desde la lógica borrosa y la lógica natural se han desarrollado algunas aproximaciones que intentan superar estas limitaciones. La lógica borrosa se han enfocado, principalmente, a incluir los cuantifica-dores vagos (Zadeh 1985; Dubois and Prade 1988) sin incrementar el número de premisas y términos manejados. La lógica natural, por su parte, se ha centrado en incrementar el número de premisas y términos usados, pero limitándose a los cuatro cuantificadores y tres términos clásicos (Sommers 1982).

Sin embargo, creemos que existen otras vías de ampliación para el silogismo. En este trabajo presentamos una expansión en la que proponemos combinar el patrón básico del silogismo con la analogía, obteniendo así un patrón de inferencia plausible. En esta versión inicial, nos limitamos a la forma del Cuadrado Lógico de la Oposición) y a modelar la analogía mediante cuantificadores de similitud en su versión más básica tomada de la lógica borrosa.

Este trabajo, que se divide en 4 secciones, se organiza del siguiente modo: en la sección 2 se presentan las características fundamentales de la inferencia plausible y la analogía; en la sección 3 proponemos nuestro modelo de silogismo encadenado aproximado y, finalmente, en la sección 4, resumimos las conclusiones del trabajo.

2 El razonamiento plausible

La lógica se ha fijado tradicionalmente en el razonamiento demostrativo. Este se caracteriza por ser seguro, definitivo y estar libre de cualquier tipo de controversia (Pólya 1966). Gracias a estas propiedades, obtenemos modelos rígidos, codificados y claros que nos permiten fundamentar sólidamente el conocimiento. La máxima expresión de este tipo de razonamiento es el razonamiento deductivo, siendo el silogismo aristotélico, como hemos dicho, uno de sus ejemplos.

Sin embargo, en nuestra vida ordinaria necesitamos emplear otras formas de razonamiento, puesto que no siempre es posible hacer deducciones: muchas veces carecemos de toda la información necesaria o tenemos simples conjeturas de las que podemos extraer nuevas conclusiones con mayor o menor credibilidad o fiabilidad, pero sin llegar a ser seguras o irrevocables. El tratamiento

de este tipo de argumentaciones también ha tenido su hueco en la filosofía y podríamos situarlas bajo una de las acepciones de 'entimema'.

En su formulación tradicional, el entimema se define como "silogismo abreviado en el que se omite una de las premisas por ser evidente o darse por sobreentendida" (Vega 2011). Sin embargo, en uno de sus significados (Vega 2011), el entimema también denota a la plausabilidad como tipo de argumentación central en la dialéctica y la retórica. Por ejemplo, si estamos en el mes de julio, el cielo está despejado y el sol brilla con fuerza, es plausible pensar que en la calle hará calor. En nuestra propuesta y a la retórica como estando presentes en nuestra argumentación ordinaria, podemos concluir que nuestro trabajo tiene como objeto distintas formas del lenguaje común.

A continuación, analizaremos el patrón básico de la inferencia plausible y la analogía como el marco de desarrollo para nuestra propuesta de silogismo plausible.

2.1. Patrón básico de inferencia plausible

G. Pólya (Pólya 1966) presenta las características fundamentales del razonamiento plausible, describiéndolo como una forma de razonamiento azarosa, discutible y provisional. En su exposición, Pólya se centra en el ámbito del razonamiento matemático, aunque es extensible a otros ámbitos.

Así pues, sean A una conjetura y B una consecuencia de A; el patrón básico de inferencia plausible permite inferir una conclusión no segura, pero sí con un cierto crédito, tal y como se muestra en la Tabla 2.

Tabla 2. Patrón básico de inferencia plausible.

A implica B
B es verdadera
A es más digna de crédito

Como se puede observar, la 'conclusión' indica la fiabilidad de la conjetura (A) a partir de la verdad de su consecuencia (B). Si en lugar de ser verdadera, B fuese casi falsa entonces concluiríamos que la verdad de A también sería menos fiable; si B fuera falsa, sería suficiente para la refutación de A.

2.2. La analogía

La analogía es una forma de razonamiento muy común en los seres humanos para esbozar una conclusión. B. Bouchon-Menier (Bouchon-Meunier and Valverde 1993), en una de sus acep-

ciones, define el razonamiento analógico como aquel que se basa en la proximidad entre dos objetos o situaciones; dadas dos proposiciones o términos, si *A* implica *B* y observamos *A' (*que es similar a *A* en algún sentido), entonces es posible inferir, por la misma relación, un *B'* similar a *B*. Este proceder sólo permite extraer una conclusión con un grado de fiabilidad o probabilidad. Un ejemplo típico de analogía en el campo de la biología molecular es la explicación del metabolismo celular usando el fuego: del mismo modo que el fuego necesita oxígeno para la combustión, la mitocondria de la célula necesita oxígeno para la transformación de la glucosa.

Actualmente, las dos teorías de la analogía más ampliamente aceptadas son (de Donato 2011, p. 48): *i)* el modelo de J. S. Mill, basado en la idea de semejanza (compartir propiedades) entre los término en comparación; *ii)* el modelo de Gentner, basado en la idea de similitud estructural, donde la comparación entre los términos de la analogía se rige por tres principios: semejanza (compartir propiedades), estructura (patrones comunes) y propósito (fin perseguido por el agente). De acuerdo con el objetivo de este trabajo, el modelo de Gentner, al ser más restrictivo en la construcción de la analogía, garantiza una mayor fiabilidad a la hora de razonar.

Pólya también define la inferencia plausible y la analogía en el ámbito de la matemática (Pólya 1966, p. 310). Así pues, dos proposiciones dadas *A* y *B* decimos que son análogas si existe una hipótesis *H* de la que tanto *A* como *B* son deducibles (ver Tabla 3, parte izquierda). Una vez establecida la analogía, Pólya aplica el patrón básico de inferencia plausible y obtiene un nuevo patrón de razonamiento (ver Tabla 3, parte derecha), donde la verdad de una de las proposiciones análogas permite concluir la credibilidad de la otra en un alto grado.

Tabla 3. Patrón básico de inferencia analógica.

$$\frac{\left.\begin{array}{l} A\ es\ consecuencia\ de\ H \\ B\ es\ consecuencia\ de\ H \end{array}\right\}\ A\ es\ análoga\ a\ B}{\begin{array}{c} B \text{ es verdadera} \\ \hline A \text{ es más digna de crédito} \end{array}}$$

Al igual que en el patrón básico de inferencia plausible, si *B* es casi falsa, *A* también es menos fiable.

3 Modelo de silogismo encadenado aproximado

Como hemos indicado, el "motor de inferencia" del silogismo es el encadenamiento de los Términos Menor y Mayor a través del Término Medio según las proposiciones de las dos premisas. Nuestra propuesta consiste en incorporar la analogía al rol de 'Término Medio' y obtener así un patrón de inferencia plausible o aproximada donde la conclusión no sería incuestionable, sino que tendría un grado de fiabilidad o credibilidad. Cuanta mayor sea la similitud entre los dos términos

que desempeñan el rol del Término Medio, mayor es la credibilidad de la conclusión. A este nuevo patrón de inferencia lo denominamos *silogismo encadenado aproximado*.

Tabla 4. Razonamiento silogístico plausible.

Todos los *aficionados a la música clásica* son melómanos
Algún *aficionado a la ópera* es gallego
Algún gallego es melómano

Consideremos el ejemplo de la Tabla 4. Como podemos observar, el rol de Término Medio es desempeñado por dos términos distintos, *aficionados a la música clásica* en la Premisa Mayor y *aficionados a la ópera* en la Premisa Menor. El *término aficionados a la música clásica* denota el conjunto cuyos elementos son todos los aficionados a las obras musicales que se califican como música 'culta', 'académica' o 'docta'. Por *aficionados a la ópera* denotamos el conjunto de obras teatrales cuyo texto se canta, total o parcialmente, con acompañamiento de orquestra. Así definidos, podemos decir que ambos términos son análogos o muy similares entre ellos, por lo que la conclusión del silogismo resulta digna de crédito, razonable o aceptable.

Consideremos ahora el ejemplo de la Tabla 5. En este caso, mantenemos el mismo Término Medio en la Premisa Mayor, pero substituimos el de la Premisa Menor, *aficionados a la ópera*, por *aficionados a los musicales*. Este denota el conjunto de gente que disfrutan con las obras musicales pertenecientes al género teatral o cinematográfico de origen angloamericano, donde la acción se desarrolla con partes cantadas y bailadas. En este caso, el solapamiento entre la música clásica y los musicales es menor que el solapamiento entre *música clásica* y *ópera*, por lo que la conclusión del silogismo ya no resulta tan plausible o aceptable.

Tabla 5. Razonamiento silogístico menos plausible.

Todos los *aficionados a la música clásica* son melómanos
Algún *aficionado a los musicales* es gallego
Algún gallego es melómano

En resumen, y también de acuerdo con Pólya, podemos afirmar que cuanto mayor sea la similitud o la analogía entre los términos que desempeñan el papel de 'Término Medio', más credibilidad tiene la conclusión. En el caso extremo en que los dos términos sean idénticos, obtenemos un silogismo deductivo o demostrativo.

Es importante señalar que, en esta aproximación a la analogía y al razonamiento plausible, la vaguedad tiene un papel fundamental. Esta no sólo está presente por naturaleza en ciertos términos

como los cuantificadores tipo *muchos, pocos, bastantes,…* sino que también la asumimos con naturalidad y con efectividad en el uso de gran número de conceptos, como aquellos con los que se ejemplifica la paradoja del sorites; e.g. *calvo* o *montón*. Ahora bien, ¿existe alguna forma de "medir" cuan similares son dos términos? ¿Existen algún recurso en el lenguaje ordinario que nos permitan abordar de un modo más o menos intuitivo esta cuestión?

En esta primera aproximación al problema, proponemos modelar explícitamente la similitud o analogía entre los Términos Medios y añadirla como una premisa más al silogismo. Para desarrollar esta idea, aplicaremos los llamados *cuantificadores de similitud*, que nos ofrecen una medida de semejanza entre ellos; es decir, nos indican cómo de parecidos o diferentes son los conjuntos sobre los que cuantifican.

Los *cuantificadores de similitud* no aparecen directamente como una partícula específica en el lenguaje natural, sino que adoptan la forma de expresiones como "*A* y *B* se parecen mucho entre ellos", "los conjuntos *A* y *B* contienen aproximadamente los mismos elementos",. . . donde *A* y *B* denotan términos de similitud basada en el cálculo de cardinalidades entre conjuntos. Para poder manejar la vaguedad e interpretar expresiones como "muy similares" o "poco similares", proponemos el uso de la lógica borrosa. En particular, adoptaremos el marco para el modelado básico de la cuantificación borrosa propuesto por D. Dubois y H. Prade (Dubois and Prade 1988) para el silogismo borroso, donde a cada cuantificador Q se asocia un intervalo $[a,b] \in [0,1]$, en el que a denota el umbral inferior y b el umbral superior.

Así pues, sea E el universo de referencia y $\mathcal{P}(E)$ el conjunto potencia de E. Sean Y_1 e Y_2 dos subconjuntos de $\mathcal{P}(E)$ y Q un cuantificador de similitud básico que depende de las cardinalidades $|Y_1 \cap Y_2|$, $|Y_1 \cup Y_2|$ con su correspondiente intervalo $[a,b]$:

$$Q: \mathcal{P}(E) \times \mathcal{P}(E) \to \{0,1\}$$

$$(Y_1, Y_2) \to Q(Y_1, Y_2) = \begin{cases} 1: Y_1 \cup Y_2 = \emptyset \\ 1: \dfrac{|Y_1 \cap Y_2|}{|Y_1 \cup Y_2|} \in [a,b] \\ 0: \dfrac{|Y_1 \cap Y_2|}{|Y_1 \cup Y_2|} \notin [a,b] \end{cases}$$

Para los casos extremos, si $(a = b) = 1$, decimos que Y_1 e Y_2 son iguales; si $(a = b) = 0$, decimos Y_1 e Y_2 son completamente diferentes.

Tabla 6. Silogismo encadenado aproximado.

Todos los *aficionados a la música clásica* son melómanos
Algún *aficionado a la ópera* es gallego
***Aficionados a la ópera* y *aficionados a la música clásica* son similares en grado [0.3,0.8]**
Algún gallego es melómano (conc una confianza o fiabilidad [0.47,1])

Con el objetivo de ejemplificar esta noción, sea la Tabla 6 una versión del ejemplo recogido en la Tabla 4. Como se puede observar, respecto de la Tabla 4 se incluye una premisa adicional en la cual, mediante un cuantificador de similitud, se representa la semejanza entre los dos conceptos que desempeñan el rol de Término Medio.

Figura 1. M denota el conjunto de los melómanos, AMC el de los *aficionados a la música clásica* y AOP el de los *aficionados a la ópera*.

(a) Límite superior de Q (b) Límite inferior de Q

Por ejemplo, en Santiago de Compostela hay unos 95.000 habitantes, de los cuales asumimos que nos 90.000 son nacidos o criados en Galicia, es decir, gallegos. Por otra parte, asumimos que en la ciudad hay 5.000 melómanos y 2/5 partes de ellos, es decir 2.000, son aficionados a la música clásica. De esos 2.000, consideremos que 4/5, 1.600, también son aficionados a la ópera de un total de 3.000.

Según las cardinalidades de los términos que desempeñan el rol de término medio, el cuantificador de similitud asociado es Q =[0.47,0.8]. Este intervalo viene determinado por la distribución de los elementos. En el caso de que todos los aficionados a la ópera sean también aficionados a la *música clásica* (ver Figura 2(a)), obtenemos el límite superior $\left(\frac{1.600}{2.000} = 0.8\right)$; en el caso opuesto (ver Figura 2(b)), es decir, sólo 1.600 aficionados a la ópera también lo son a la música clásica y el resto no, obtenemos el umbral inferior $\left(\frac{1.600}{1.600+400+1.400} = 0.47\right)$.

Dada la relación definida entre los términos que desempeñan el rol de término medio, para el umbral superior (0.8) podemos asignar un grado de credibilidad total a la conclusión (1), puesto que cualquier gallego aficionado a la ópera, también será melómano. Para el umbral inferior (0.47), significa que aproximadamente la mitad de los que son aficionados a la ópera no lo son a la música clásica, por lo que no podemos concluir si son melómanos o no; por lo tanto, no podemos darle a la conclusión más que un grado de credibilidad medio. En resumen, dados los valores credibilidad obtenidos para la conclusión, podemos dar esta argumentación por bastante plausible.

4 Conclusiones

En este trabajo hemos realizado una exploración inicial de una expansión del silogismo clásico hacia el razonamiento plausible o aproximado. También hemos introducido el concepto de analogía en el rol del Término Medio como factor determinante de la credibilidad de la conclusión: a mayor grado de similitud entre los conceptos del Término Medio, mayor credibilidad de la conclusión. La similitud se puede hallar usando aproximaciones borrosas.

Como trabajo futuro, proponemos seguir profundizando en el desarrollo de este modelo, así como analizar la posibilidad de desarrollar una medida de similitud basada en la sinonimia de las palabras, lo que nos permitiría dar un paso hacia una forma de silogismo plausible de tipo más lingüístico. Proponemos también el estudio de nuevos patrones de silogismo que incluyan a los cuantificadores borrosos (*muchos, pocos, bastantes,…)*, además de los cuatro clásicos y que manejen un mayor número de términos y de premisas.

Agradecimientos

Este artículo fue realizado con el apoyo Ministerio de Economía e Innovación y el Fondo Europeo de Desarrollo Regional (ERDF/FEDER) (Proyecto TIN2011-29827-C02-02) y del Ministerio de Educación, Cultura y Deporte (Becas del Programa FPU).

REFERENCIAS

Aristóteles. 1982-1986. *Organon*. Madrid: Gredos. Bouchon-Meunier, B., and L. Valverde. 1993. “Analogy relations and inference.” *Second IEEE International Conference on Fuzzy Systems, 1993.* 1140–1144.

Brown, M. 1984. “Generalized Quantifiers and the Square of Opposition.” *Notre Dame Journal of Formal Logic* 25 (4): 303–322.

Corcoran, J. 2011. Deducción / Deducibilidad en *Compendio de lógica, argumentación y retórica*, L. Vega y P. Olmos (eds.), 168–170. Madrid: Ed. Trotta. de Donato, X. 2011. Analogía en Compendio de lógica, retórica y argumentación, L. Vega y P. Olmos (eds.), 47–50. Madrid: Ed. Trotta.

Dubois, D., and H. Prade. 1988. "On fuzzy syllogisms." *Computational Intelligence* 4 (2): 171–179. Pólya, G. 1966. *Matemáticas y razonamiento plausible*. Madrid: Ed. Tecnos.

Pólya, G. 1966. *Matemáticas y razonamiento plausible*. Madrid: Tecnos.

Sommers, F. 1982. *The Logic of Natural Language*. Oxford: Clarendon Press.

Vega, L. 2011. Entimema en Compendio de lógica, retórica y argumentación, L. Vega y P. Olmos (eds), 226–228. Madrid: Ed. Trotta.

Zadeh, L. A. 1985. "Syllogistic reasoning in fuzzy logic and its applications to usuality and reasoning with dispositions." *IEEE Transactions On Systems, Man and Cybernetics* 15 (6): 754–765.

CAUSALIDAD CON REDES DE PETRI

A. SOBRINO
USC

ABSTRACT: In this paper I deal with causality using Petri Nets, exemplifying necessary and sufficient causal conditions, causal relations, and some of their most common varieties, as preemption, trumping, fizzling or overlapping.

Key words: Causality, Petri Nets, causal relation, preemption, trumping, fizzling, overlapping.

1 Introducción. Dos interpretaciones de la causalidad

Hay cuando menos dos interpretaciones habituales de la causalidad:

1. Causalidad como una relación formal entre causa y efecto. En ese caso, la causa es suficiente, necesaria o simplemente incrementa la probabilidad del efecto.
2. Causalidad como la transmisión de alguna marca de la causa al efecto.

A la primera modalidad se la conoce como causalidad condicional; a la segunda como causalidad material.

La causalidad condicional se enuncia en términos de causas suficientes o necesarias y se expresa usualmente así:

c causa *e* si tanto *c* como *e* ocurren y *e* ocurre si *c* ocurre

Es decir, *c* es suficiente para provocar *e*. En el caso de que *c* sea necesario, la causalidad se formula mejor usando condicionales contrafácticos. Así, se dice que *c* es necesario para *e* en el caso de que si *c* no hubiese ocurrido, tampoco hubiese sucedido *e*; es decir, *c* es imprescindible para *e*. Las causas suficientes pueden ser deterministas o simplemente favorecer al efecto. En este caso se dice que la $P(e|c) > P(e)$; es decir, que la probabilidad del efecto dada la causa es mayor que la probabilidad del efecto solo (sin la causa).

La causalidad material o física entiende el proceso causal no (sólo) en términos de causas suficientes o necesarias, sino como algún trasvase de la causa al efecto. En este caso, atiende a un esquema como el siguiente:

c causa *e* si se transmite una marca (un átomo de materia, un *quantum* de energía o un *bit* de información) de *c* a *e*

Las concepciones condicional y material de la causalidad no son antitéticas. Sin ir más lejos, la versión material se formula de manera condicional. La tesis de este trabajo es que un tipo de autómatas, las Redes de Petri (sencillas (RP) y temporizadas (RPT)), reflejan bien esta dualidad y sirven para ejemplificar las relaciones causales en términos condicionales y materiales. Nuestra propuesta se inspira, aunque no coincide, con la sugerencia de Lewis (Lewis 1986), de representar los procesos causales con 'neurores'.

Las RP se han mostrado útiles para modelar procesos donde hay concurrencia de eventos y flujo de información, situación típica de las relaciones pluricausales. Una RP (Murata 1989) es un grafo dirigido que consta de los siguientes componentes:

- Marcas: representan recursos en un sentido amplio: pueden ser tanto recursos físicos como inmateriales (energía, información, etc.)
- Lugares: son los sitios donde se almacenan los recursos a la espera de ser trasvasados.
- Transiciones: habilitan la transición de un lugar a otro, conectando los arcos lugar-transición y transición-lugar. Pueden tener pesos, que se interpretan así:
    - Pesos de los arcos lugar-transición: representan el número mínimo de recursos que debe tener el lugar para que se lleve a cabo la acción representada por la transición.
    - Pesos de los arcos transición-lugar: representan el número de recursos que es almacenado en ese lugar y que son producidos al ejecutarse la transición.

Los arcos pueden ser excitadores (→) o inhibidores (→o). A las transiciones se les pueden asignar probabilidades, que indican la frecuencia con que son disparadas cuando se satisfacen las condiciones oportunas.

Las RPT son como las RP sencillas, pero añaden una posible temporización tanto en los lugares como en las transiciones. Una anotación temporal en los lugares se interpreta como el tiempo mínimo que debe permanecer en el mismo una marca antes de habilitar el disparo de una transición posterior a dicho lugar; una cifra en las transiciones se lee como el tiempo que separa el comienzo (el consumo de marcas en lugares anteriores) y la finalización del disparo de la acción (la producción de marcas en lugares posteriores).

Gráficamente:

Ejemplo de RP:

Esta RP representa un posible escenario de bebidas, consumidores y camareros, en el que la causa de que sirvan una bebida es que haya existencias (X) y el camarero (M) esté libre. En el momento de servir, M consume su marca; es decir, no puede servir a otro cliente. Una vez atendido, el camarero queda de nuevo libre (M tiene una marca) y el proceso se reinicia. Así hasta que se consumen las tres bebidas.

2 Causalidad condicional

Las causas suficientes determinan el efecto. Veamos a continuación la causalidad determinista en términos de causas suficientes y necesarias, su formalización en términos condicionales y su ejemplificación usando Redes de Petri. Hay cuatro escenarios posibles:

Caso	C es necesaria	C es suficiente
1	+	+
2	+	-
3	-	+
4	-	-

Ejemplifiquémoslos en términos de la relación causal que hay entre una serie de factores F={picadura mosquito anofeles, bacteria *Bordetella pertussis*, bacteria *Bordetella parapertussis*, *Actinobacillus lignieresii,* fumar cigarrillos, no usar filtros, propensión genética, inhalar gas radón, prescindir de mascarilla) y un conjunto de enfermedades E={malaria, tosferina, infección ovina, cáncer de pulmón}.

Caso_1. A es suficiente y necesaria para E.

- Ejemplo: La picadura del mosquito anofeles (A) causa malaria (E).
- Formalización: A → E
- Interpretación: Un efecto es producido por una única causa. Si A está presente, E se produce y, por ser causa única, si ocurre E es que se ha producido A, pero no se contempla A ↔ E, debido a la restricción a la que obliga una ley básica de la causalidad: A debe ser anterior a E).
- RP:

Es decir, A está marcada y habilita la transición de su marca a E.

Caso_2. A es una causa necesaria, pero no suficiente para C

- Ejemplo: Es necesaria la presencia del germen *Actinobacillus lignieresii* (A) para que se produzca una infección en el ganado (bovino y ovino) (E); pero sólo se desarrolla la enfermedad si hay otros factores coadyuvantes que dañen la mucosa (B) (p. ej., comer vegetales puntiagudos o abrasivos). En ausencia de ese daño, la bacteria puede estar presente y la enfermedad no producirse.
- Formalización: *Si A∧B entonces E.*
- Interpretación: A es necesario, pero no suficiente para E.
- RP:

Dado que A y B tienen marca, ambas habilitan la transición a E. Para que se dispare la transición no basta con que A tenga una marca; también la requiere B. A es necesaria, pero no suficiente para causar E.

Caso_3. A es una causa suficiente, pero no necesaria para E.

- Ejemplo: Es suficiente tener bacteria *Bordetella pertussis* (A) para desarrollar la tosferina. Pero no es necesario ser infectado con esa bacteria. También puede ocasionar esa enfermedad la bacteria *Bordetella parapertussis* (B).
- Formalización: *Si* $A \vee B$ *entonces E*.
- Interpretación: A es suficiente, pero no necesario para E.
- RP:

Es decir, tanto A como B contienen marcas que habilitan, por separado, la transición que lleva al marcado de E.

Caso_ 4. A no es ni suficiente ni necesaria para E.

- Ejemplo: Tener presbicia (A) no es causa suficiente ni necesaria de padecer un infarto de miocardio (E).
- Formalización: B → E. A. (B es, p. ej., tabaquismo)
- Interpretación: Una causa que no es necesaria ni suficiente es aquella que no muestra relación alguna con lo causado.
- RP:

E depende de B, no de A, que le es ajena.

Hemos visto algunos casos en los que se contempla una causa limitada a uno o dos factores causales. Pero las situaciones que involucran a más de dos factores son muy habituales. La causalidad *INUS* propuesta por Mackie, (1980), hace referencia a un marco en el que confluyen varias causas complejas y donde un factor causal sencillo es interpretado como una parte insuficiente, pero no

redundante, de una parte no necesaria, pero suficiente, para provocar el efecto (***I****nssuficient but* ***N****ot redundant part of a* ***U****nnecessary but* ***S****ufficient cause*, ***INUS***).

- Ejemplo: El cáncer de pulmón (H) puede ser producido quizás por inhalar gas radón (D) y prescindir de mascarilla (F) o por inhalar humo de tabaco (A), no usar filtros (B) y tener predisposición genética (C).
- Formalización: *Si* $(A \wedge B \wedge C) \vee (D \wedge F) \rightarrow H$.
- Interpretación: El grupo $(D \wedge F)$ no es necesario, sino suficiente para causar H, e incluye factores (D) que son necesarios, aunque no suficientes, para causar H.
- RP:

La causalidad INUS expresa de modo paradigmático la interrelación que hay entre causas necesarias y suficientes en las situaciones causales.

A veces la causa provoca inequívocamente el efecto *(Ingerir 2 gr. de cianuro causa la muerte)* y, a veces, simplemente lo favorece *(Beber alcohol provoca a menudo accidentes de carretera).* Cuando una causa favorece un efecto se dice que las causas incrementan la probabilidad del efecto. Formalmente:

c causa *e* syss $P(e \mid \neg c) < P(e \mid c)$

En las RP, la probabilidad puede ser expresada anotando la frecuencia de disparo de una regla una vez satisfechas las condiciones usuales. Así,

indica que la transición se dispara 6 de cada 10 veces.

A la idea de que la causa incrementa la probabilidad del efecto se le han puesto algunas objeciones. A continuación, expondremos algunas de las más usuales y las ilustraremos usando RP. Se dividen fundamentalmente en dos apartados:

1) Causas que lo son, pero que no incrementan la probabilidad del efecto.
2) Factores que incrementan la probabilidad del efecto sin ser causa del mismo.

2.1. Causas que lo son, pero que no incrementan la probabilidad del efecto.

2.1.1. *Preemption* (prevención)

La prevención dice que, en una situación pluricausal, no siempre la causa eficiente es la que más probabilidad le aporta a la ocurrencia del efecto.

- Ejemplo: Ana (A), una chica sensata, en un momento de locura arroja una piedra contra una ventana de cristal. Pedro (P), un reconocido gamberro, también tiene una piedra en su mano, pero al ver que Ana arroja la suya, desiste. La acción de Ana inhibe a Pedro. Por su reconocido gamberrismo, la presencia de Pedro le otorga más probabilidad de romperse al cristal que la presencia de Ana; sin embargo, es Ana la causa y no Pedro. Moraleja: quien más probabilidades tiene de producir el efecto no es el que lo provoca.
- RP:

A inhibe la acción de P; aún cuando este último tiene más probabilidad de romper el cristal, es A finalmente quien ocasiona su rotura.

2.1.2. *Trumping* (superación)

La superación dice que aún teniendo dos factores con la misma probabilidad de producir el efecto, el que ocurre antes, no inhibe, sino que supera al otro, a la hora de acreditarse como su causa. El factor que provoca la causa es suficiente, pero no necesario, para el efecto.

- Ejemplo: En una tierra mágica hay dos brujos, Merlín *(Me*) y Morgan (*Mo*). Cada uno de ellos puede arrojar un hechizo sobre el príncipe convirtiéndolo, a media noche, en rana. Las leyes de la magia dicen que el primer hechizo que se produzca durante el día será el que ocasiona la acción. Y sucede que Merlín arroja su hechizo por la mañana y Morgan por la

tarde. A media noche el príncipe se convierte en rana y no hay duda de que fue el hechizo de Merlín la causa. El hechizo de Merlín no desactivó al de Morgan; si uno de ellos no hubiera tenido efecto, lo habría tenido el otro. Aunque los dos tienen la misma probabilidad de provocar el efecto, Merlín supera a Morgan por haber obrado antes. Merlín es el causante, pero no por conferir más probabilidades que Morgan al efecto. Moraleja: aún teniendo probabilidades iguales, el que se adelanta, causa el efecto.

- RPT:

Como se observa, el nodo *Mo* tiene un retardo 5, mientras *Me* tiene retardo 1. Por tanto, se habilita antes la transición de *Me* para conseguir el efecto deseado.

2.2. Incremento de la probabilidad del efecto sin ser su causa.

2.2.1. *Fizzling* (fracaso)

Hay factores que, favoreciendo o afectando a la probabilidad del un evento, fracasan en ser su causa.

- Ejemplo. Pedro coge una piedra y Ana, ajena a lo que hace Pedro, arroja su piedra contra la ventana (E). Ana se adelanta a Pedro que, despistado, mira para otro lado y no se entera. Supongamos que en el momento en el que Pedro coge la piedra hay una probabilidad distinta de 1 de que Ana impacte contra el cristal y una probabilidad diferente de 0 de que Pedro haga lo propio. La disposición fracasada de Pedro no es la causa, obviamente, de que el cristal se haya roto, pero su presencia incrementa -dada la probabilidad distinta de 0- la posibilidad de tal suceso. Moraleja: a pesar del fracaso, la presencia de Pedro favorece al efecto, pero no es suficiente para su causación.
- RPT:

Obsérvese que mientras el retardo de P es 5, el de A es 1. Sin embargo, la probabilidad de que se dispare t_1 es 0,6 y la probabilidad de que se dispare t_2 es 0,4. Multiplicando ambos factores resulta que $(A, t_2) > (P, t_1)$. Por tanto, A es la causa de E.

2.2.2. *Overlapping* (solape)

A veces dos causas contribuyen, en principio, con la misma probabilidad a un efecto y si bien, finalmente, sólo una de ellas es la causa fetén, la presencia de la otra causa contribuye, en cualquier caso, a incrementar la probabilidad del efecto.

- Ejemplo: Como en el ejemplo de *Trumping*, en este escenario hay dos magos involucrados y tres personajes nuevos: el rey (R), el príncipe (P) y la reina (Ra). Supongamos que el hechizo de Merlín tiene una probabilidad 0.5 de convertir al rey y al príncipe en rana y que el hechizo de Morgan, independiente del de Merlín, tiene una probabilidad 0.5 de convertir al príncipe y a la reina en rana. Finalmente, son rey y el príncipe, pero no la reina, los que se convierten en ranas. Parece que el hechizo de Morgan no es la causa que el príncipe se hubiera convertido en rana (el hecho de que la reina no haya sufrido el hechizo así lo muestra). Pero es el caso que Morgan, aun no siendo causa del mismo, incrementa la probabilidad del hechizo del Príncipe. Moraleja: un efecto puede ser debido a dos causas igualmente probables y una de ellas, aun no siendo la causa fetén, incrementa la probabilidad del efecto.
- RP:

En conclusión, la prevención o el fracaso, en sus versiones estándar, refinada o ampliada, son incapaces de proporcionar condiciones necesarias o suficientes para la causación.

3 Causalidad material

Algunos filósofos han rechazado una aproximación a la causalidad en términos de condiciones necesarias o suficientes y se han mostrado a favor de una interpretación de la relación causal como un proceso en el que hay conexiones físicas. Así, (Fair 1979), identifica la causación como flujo de energía; (Salmon 1984), como la transmisión de una marca (la capacidad de propagar su propia estructura) y (Ehring 1997), como la persistencia de un tropo de la causa al efecto.

A la causalidad como mecanismo que transmite algo se le han puesto algunas objeciones. Como antes, trataremos de clarificarlas usando las redes de Petri. Veamos primero un ejemplo de causalidad material con Redes de Petri y después algunos de los problemas puestos.

Esta red modela una reacción química: la combinación de dos moléculas de hidrógeno y una de oxígeno produce una molécula de agua.

A continuación, describimos algunos problemas asociados a esta interpretación de la causalidad:

3.1. *Disconnections* (desconexiones)

Puede haber factores que ejerzan una influencia causal sobre el efecto sin formar parte de un mecanismo que conduce a él.

- Ejemplo: Un avión (A) sigue una dirección que le lleva a colisionar con una montaña. El controlador (C) alerta al piloto del tal situación, indicándole que haga un giro de rumbo. Un saboteador (S) destruye en ese momento la emisora de radio, de manera que la alerta no llega al piloto y el avión se estrella.
- RP:

No hay duda de que el sabotaje es la causa del accidente. Pero no hay un proceso que conecte al sabotaje con el impacto del avión. No hay una continuidad entre esos nodos. Conclusión: no es necesaria una transmisión de marcas para la causación.

3.2. Trazas (Traces)

Hay acciones que tienen un enlace directo con el efecto, pero no son su causa.

- Ejemplo: Ana arroja un ladrillo contra una ventana mientras Pedro hace lo mismo. Pero alguien involuntariamente le toca en el codo a Pedro en el momento de lanzar, de manera que su piedra describe una parábola y resbala por la superficie del cristal sin provocarle daño alguno. Aun habiendo un trazo o conexión entre la acción de Pedro y el impacto en el cristal, no es causa de su rotura.
- RP:

4 Conclusión

Las RP sencillas o temporizadas ilustran de modo adecuado tanto una interpretación formal como una interpretación material de la causalidad, ayudando a distinguir los diferentes tipos de causas, así como las diversas dificultades que, en las distintas interpretaciones, se le han puesto.

Agradecimientos

Este artículo fue realizado con el apoyo del Ministerio de Economía e Innovación y el Fondo Europeo de Desarrollo Regional (ERDF/FEDER) (Proyecto TIN2011-29827-C02-02).

REFERENCIAS

Ehring, D., 1997, *Causation and Persistence*. N. Y., Oxford**.**
Fair, D., 1979, 'Causation and the Flow of Energy', *Erkenntnis*, XLVIII, 219-250
Lewis, D., 1976, 'Postscripts to 'causation'', in *Philosophical Papers*, vol. 2, Oxford University Press, 172-213.
Mackie, J. L., 1980, *The Cement of the Universe: A Study of Causation*, Oxford University Press.
Murata, T., 1989, 'Petri Nets: Properties, Analysis and Applications', *Proceedings of the IEEE*, vol. 77, nº 4, April, 541-580.
Salmon, W., 1984, *Scientific Explanation and the Causal Structure of the World*. Princenton University Press.
Schaffer, J., 2001, 'Causes as probability-raisers of process', *Journal of Philosophy*, 75-92.

TEORÍA DE LA ARGUMENTACIÓN: PERSPECTIVAS Y PROBLEMAS*

Coordinador LUIS VEGA REÑON
UNED

En un campo que está experimentando un crecimiento exponencial de propuestas y contribuciones, como es hoy el de los estudios en torno a la Argumentación, conviene una revisión periódica no solo de los desarrollos más o menos consolidados, sino de los problemas pendientes y las nuevas perspectivas. Este simposio puede representar una brújula y un incentivo para los interesados en las ideas en curso sobre la argumentación y para los preocupados por sus prácticas en nuestro contexto.

EXPERTOS Y AUTORIDADES EN LA PRÁCTICA ARGUMENTATIVA

BEGOÑA CARRASCAL[1]
Departamento de Lógica y Filosofía de la Ciencia
Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)
b.carrascal@ehu.es

ABSTRACT: The appeal to the authority of experts is nowadays difficult to avoid when public decisions about complex matters have to be taken. This kind of argumentation is also widely used to support our academic papers. Several critical questions are usually associated with the so called 'appeal to expert opinion' argumentation scheme in order to critically asses the validity or the cogency of it. Those questions, if raised by an opponent in a discussion, have the property of shifting the burden of proof so that the arguer may repair or withdraw her proposal.

But looking at the argumentations as communicative acts we realize that this is not the usual reaction when presented with an 'appeal to expert opinion' in many discussions. Communication goes fast and to critically assess everything that is said would be uneconomical and implausible. Besides, in many communicational settings we are in no position of questioning the authority because we don't have the necessary knowledge about the matter and, also, because, maybe, the claim proposed by the authority is coherent with our system of beliefs. Even in those cases in which we don't accept the claim, instead of questioning the expert, we may look for another expert in order to support an alternative option. In this work we present several examples of the inadequacy of the

* Desarrollado en el marco del proyecto financiado por la Secretaría de Estado de Investigación, desarrollo e Innovación, FFI2011-23125.
[1] Ponencia parcialmente financiada por el proyecto FFI 2008-20118 del Ministerio de Ciencia e Innovación.

standard list of questions linked to the 'appeal to expert opinion' argumentation scheme to conclude that any descriptive enough model of argumentative practice has to include elements of the communicative context and of the audience. The usual analyst's point of view doesn't explain how we deal with real arguments in practice, because ordinary argumentations have to be assessed in a definite social setting in order to improve them.
PALABRAS CLAVE: Argumentación, comunicación, esquema argumentativo, apelación al experto

La cantidad de información a nuestra disposición hoy en día sobre prácticamente cualquier tema, se ha incrementado enormemente debido al acceso a las nuevas tecnologías de la información. Sin embargo, esa misma cantidad de información y el gran desarrollo científico y tecnológico de los últimos tiempos hace que no siempre sea posible entender, ordenar y elegir de entre todos los contenidos a nuestra disposición aquello que es necesario para poder decidir sobre una cuestión directa y autónomamente. En consecuencia, la apelación a la autoridad de los expertos es una práctica habitual a la hora de tomar decisiones sobre muchos aspectos de la vida pública. Es también una forma común de apoyar nuestros trabajos en el ámbito académico.

Cuando se utiliza en forma argumentativa la apelación a la autoridad se suele considerar como una instancia del esquema 'apelación al experto' ya que es usual considerar que, para que el argumento pueda ser calificado de aceptable, el tipo de autoridad debe ser de tipo epistémico y ligado a un ámbito de conocimiento en el que la autoridad sea experta. Este tipo de argumento ha estado vinculado clásicamente con la falacia denominada *ad verecundiam.*

Hay varias formas del esquema, con presuposiciones más o menos fuertes sobre la forma de inferencia o la verdad o aceptabilidad de las premisas implícitas en el esquema. Consideraremos aquí una forma débil que Walton (1997) cataloga como presuntiva y que se puede expresar de la siguiente manera:

Un experto E en un dominio D afirma A (dentro del dominio D), luego A es presumiblemente verdadera.

La premisa implícita presupuesta sería "*Lo afirmado por el experto E sobre aspectos dentro de su campo de conocimiento es presumiblemente verdadero"*.

Posteriormente a Hamblin (1970) es habitual asociar una serie de preguntas críticas con cada esquema argumentativo que pueden ser formuladas por un oponente en la discusión con el objetivo de asegurar la solidez y razonabilidad del argumento. Al formularlas, la carga de la prueba automáticamente vuelve al proponente para que este refuerce, modifique o retire la argumentación propuesta. En el caso de la apelación al experto las preguntas habituales se refieren tanto a la evaluación del contenido de la afirmación del experto (si está basada en hechos o si la afirmación del proponente coincide con lo dicho por el experto, por ejemplo), como a la del experto en sí (su reputación y credibilidad en relación a su trayectoria en un campo, consistencia con otras fuentes etc.)

La lista estándar de preguntas puede completarse con sub-preguntas más concretas si se consideren necesarias para asegurar suficientemente el contenido de la afirmación del experto. Los ejemplos de la literatura (Walton, 1997; Walton et al. 2008) solo contemplan aclaraciones de las preguntas principales, sin incluir adaptación específica a la situación argumentativa, ni referencias a la audiencia.

En los últimos tiempos investigadores tanto del área de la argumentación (Zenker, 2011; Goodwin, 2011; Carrascal, 2011) como del área de sociología de la ciencia (Collins & Weinel, 2011) han expresado sus dudas sobre la adecuación de todas o algunas de las preguntas de la lista estándar. Esta lista cuyo origen es el estudio de las falacias, está pensada más para un análisis *a posteriori* de las argumentaciones y, en todo caso, resulta útil para sacar a la luz los casos más claros de *ad verecundiam*, pero no es adecuada para mejorar la práctica argumentativa ordinaria por diferentes razones, entre ellas que no constituye una lista necesaria y suficiente de preguntas, ni describe realísticamente la forma ordinaria de reaccionar ante afirmaciones apoyadas por expertos, por no incluir aspectos contextuales ni específicos de la situación argumentativa.

Por citar algunos ejemplos que apoyen lo anterior, consideremos el ámbito de actuación pública en el que la apelación al experto es usual para avalar tomas de decisiones sobre asuntos complejos y técnicos. Es difícil saber si un experto tiene intereses o no a la hora de afirmar algo relativo a su campo, aunque se pueda sospechar que si los tiene, si el debate es intenso y el informe se presenta por una parte muy implicada en la cuestión, pero es mucho más difícil que en una discusión se cuestione sobre estos presuntos intereses y no sobre el informe presentado. Al fin y al cabo, todo el mundo tiene intereses más o menos marcados sobre cuestiones públicas existiendo múltiples ejemplos de informes técnicos contradictorios que avalan tomas de decisiones complejas sobre un mismo asunto. Si no hay hechos y datos para poder demostrar una afirmación particular (y aun en esos casos, las interpretaciones de esos datos pueden diferir) o si los intereses del supuesto experto no son muy evidentes, es difícil pensar tanto en la formulación de esta pregunta como en la existencia de una respuesta conclusiva a ella que sirva para avanzar en la discusión. La falta de consenso se podría explicar tanto por la misma complejidad del asunto tratado como por las diferentes interpretaciones del problema debido a intereses e ideologías contrapuestas. En ambos casos, no parece que la denuncia de existencia de intereses particulares del experto en la elaboración de un informe, salvo en casos muy extremos, sea habitual o sirva para avanzar en la toma de una mejor decisión.

Otro ejemplo que podemos presentar viene de la observación de la bibliografía utilizada en diferentes trabajos académicos principalmente en el campo de las humanidades. La gran cantidad de artículos sobre casi todos los temas hace que la mayoría de ellos ni sean leídos ni, por tanto se citen

(Willard, 1990). La elección de las citas no depende exclusivamente de valores epistémicos sino también de diversos aspectos sociales, entre otros, facilidad de acceso a las revistas más conocidas, contactos con investigadores reconocidos, buen dominio del inglés etc. No hay demasiados estudios empíricos sobre el tema pero simplemente observando las ponencias presentadas a una conferencia lo suficientemente amplia, es fácil constatar que entre autores americanos hay prevalencia de citas de otros autores americanos, aun en casos en los que las mismas o similares consideraciones sean hechas por autoras más noveles y/o no pertenecientes al mundo académico anglosajón. Por otra parte, el supuesto consenso por el que se puede inquirir al formular una de las preguntas críticas, no es ni tan siquiera esperado entre los diferentes autores de muchos campos, especialmente en el de las humanidades. Las discusiones se alargan interminablemente a lo largo de la historia y los consensos, si se consiguen, necesitan tiempo y se circunscriben, a menudo, a aspectos puntuales o a una época y a una comunidad epistémica (más o menos extensa).

Otra de las cuestiones se refiere a si la afirmación del experto está basada en evidencia. Esta pregunta, adecuada, por ejemplo, para examinar una toma de decisiones sobre asuntos basados en datos o en hechos (que por otro lado se han querido evitar por medio de la apelación al experto) en el caso de las apelaciones a la autoridad en las citas académicas en el campo de las humanidades, nos llevaría a través de una cadena de citaciones que eventualmente podría convertirse en un argumento circular.

Una lista de preguntas críticas como la que hoy en día se encuentra en la literatura ha sido elaborada teniendo en mente los casos más claramente falaces de aplicación del argumento de autoridad, casos en los que la autoridad no es en realidad experta y se sustenta de forma autoritaria basándose mayormente en su posición social y no en su posición como agente epistémico. Pero entre argumentos válidos y/o correctos y falacias se puede dibujar toda una línea continua de argumentaciones más o menos débiles pero que resultan aceptables en un contexto argumentativo determinado y creemos que estos argumentos se deberían analizar, reforzar o mejorar, teniendo en cuenta elementos de dicho contexto, entre ellos la audiencia a la que van dirigidos.

Si queremos llegar a entender mejor como es la práctica argumentativa ordinaria, tenemos que considerar la argumentaciones cotidianas como actos comunicativos que tienen lugar y se desarrollan en una situación comunicativa determinada, que están pensados para convencer a un interlocutor específico y que deberían ser evaluados teniendo en cuenta aspectos particulares de dicha situación. Considerar aceptable una argumentación en un contexto concreto no significa que lo sea en toda situación. En general, las argumentaciones ordinarias son débiles y presuntivas y sirven para llegar a consensos puntuales que nos permiten avanzar a la hora de tomar decisiones, pero salvo en casos en los que es posible una resolución total para demostrar su aceptabilidad, la

mayoría de ellos se sitúa en entornos de incertidumbre con conclusiones provisionales y revisables. En situaciones argumentativas cotidianas, igualmente que en otras situaciones comunicativas, un oyente incorpora progresiva y dinámicamente nuevas creencias a su sistema con criterios principalmente basados en una maximización de la coherencia, sabiendo que, en muchos casos, estas creencias serán revisadas en un futuro.

No es posible examinar exhaustivamente toda y cada una de las afirmaciones hechas por un participante en un acto comunicativo. La comunicación es rápida y los recursos limitados. Hay muchas informaciones que son compatibles (o no manifiestamente incompatibles) con las creencias que teníamos de antemano y, en general, no hay tiempo ni sería razonable examinar exhaustivamente los argumentos esgrimidos para apoyar una tesis porque ello supondría un esfuerzo excesivo que, además, impediría una comunicación efectiva.

Consideramos junto con investigadores del área de la comunicación (Sperber et al. 2010) y del área de la epistemología (Adler, 2006) que es razonable considerar que a la hora de incorporar información a nuestro sistema por medio de un acto comunicativo, se aplique, por defecto, como regla comunicativa básica que en un entorno comunicativo determinado y siempre que no haya indicaciones para lo contrario, se acepte o no se cuestione lo que se nos comunica, si ello no es contradictorio con nuestro sistema actual de creencias en ese momento comunicativo.

Esta regla explicaría por qué, en general, en muchos casos no reaccionamos críticamente a una tesis apoyada por una apelación a un experto: no tenemos información suficiente sobre ese tema, el experto está avalado por una determinada institución, dudar en dicho contexto es costoso (porque por ejemplo supone contradecir creencias arraigadas en la comunidad, etc.) No se trata de aceptar lo afirmado por el experto como una verdad absoluta, sino de una forma débil, sujeta a revisión y solo en el caso de no existir indicaciones contrarias al respecto. Si el modelo ha de ser lo suficientemente descriptivo de las prácticas argumentativas, creemos que esta regla debería ser parte del mismo.

¿Es, sin embargo, esta actitud epistémicamente razonable? ¿Se puede considerar aceptable desde un punto de vista normativo de las prácticas argumentativas o nos lleva a aceptar de forma ciega y acrítica lo afirmado por los expertos?

Experimentos realizados en el campo de la psicología cognitiva demuestran que una actitud de confianza (o de no desconfianza) hacia el otro no significa que se acepte ciegamente todo lo este que dice, ya que en todo acto comunicativo está activada una actitud de vigilancia epistémica (Sperber et al. 2011) que nos hace reaccionar críticamente si la fuente no es lo suficientemente fiable o el contenido tiene consecuencias importantes en nuestro sistema de creencias (para ese momento de la comunicación). La nueva información se añade a la que ya se compartía de

antemano entre los participantes en la conversación, y se contrasta la coherencia del nuevo sistema resultante de incorporar la nueva información recibida. En entornos argumentativos entre especialistas de un mismo campo, la verificación y control de la coherencia pueden ser, en general, más exhaustivos, sobre todo si las afirmaciones presentadas se alejan de lo establecido como verdadero en ese campo. Si la comunicación y apelación al experto se hace entre agentes próximos, los mecanismos de vigilancia pueden estar más relajados, sobre todo si la no aceptación de lo apoyado por la autoridad tiene consecuencias fuertes para la audiencia como miembros de la comunidad a la que pertenecen (entornos educativos tradicionales, militantes de una tendencia política o religiosa...). Algunos investigadores (Daukas, 2006; Bondy, 2010) ven en estos vínculos (o falta de ellos) la posibilidad de cometer "una injusticia argumentativa" ya que todos tenemos favoritismos o prejuicios hacia los miembros de las diferentes grupos a los que pertenecemos (o no) y proponen incrementar la vigilancia para evitar caer en estereotipos culturales injustos. Hay que recordar aquí que estamos hablando de contexto argumentativos, no solo de transmisión de ideas mediante comunicación. En contextos de este tipo, partimos de que si hay argumentación es porque ha habido duda o algún punto ha sido cuestionado, por tanto, creemos que es precisamente la práctica argumentativa la que puede romper el círculo para solventar los problemas planteados. Para mejorar las técnicas argumentativas, la educación, un conocimiento mayor sobre el tema y continuo debate son factores importantes que permiten desarrollar mejores mecanismos de control y de pensamiento crítico.

Por otra parte, tal y como hemos señalado, la reacción ante un argumento de apelación al experto no es siempre en el sentido propuesto por las preguntas críticas, ya que en muchos casos, en vez de cuestionar al experto o buscar una contradicción en lo que afirma, en la práctica, se buscan argumentos alternativos para apoyar un punto de vista diferente. Si hay voluntad de acuerdo, la discusión puede avanzar adoptando por ejemplo puntos intermedios aceptables para ambos participantes. El desacuerdo entre expertos no siempre desautoriza un argumento ya que no todo el mundo considera aceptable las mismas razones. Tiempo y debate continuo es necesario para llegar a aceptar una determinada interpretación de un hecho o para adoptar criterios y procedimientos aceptables sobre un punto en conflicto. Estos acuerdos son a su vez revisables a lo largo del tiempo, siendo este un proceso social en el que no solo influyen aspectos relativos al conocimiento, sino también los pensamientos de los líderes de la época y su prevalencia a lo largo del tiempo. Olmos (2008) lo expresa como parte de una dinámica social en las que lo dicho por otros es utilizado para apoyar argumentaciones que, a su vez, serán utilizadas por otros para apoyar nuevas ideas.

En conclusión, creemos que la práctica argumentativa debe ser considerada como un acto comunicativo dinámico y sujeto, por tanto, a reglas generales de comunicación. Si se quiere que

resulte en una mejora de las práctica argumentativas reales, la evaluación del argumento de apelación al experto debe incluir aspectos relativos a la situación argumentativa en el que se presenta el argumento, sin olvidar a la audiencia.

REFERENCIAS

Adler, J. 2006, 'Epistemological Problems of Testimony'. In: Zalta, E.N. (ed.). *The Stanford Encyclopedia of Philosophy* (Winter 2010 Edition), http://plato.stanford.edu/archives/win2010/entries/testimony-episprob/

Bondy. 2010, 'Argumentative Injustice'. *Informal Logic,* Vol. 30-3, 263-278.

Carrascal, B. 2011, 'Authority arguments in academic contexts in social studies and humanities.' In: Zenker, F. (ed.). *Argumentation: Cognition and Community. Proceedings of the 9th International OSSA Conference*, Windsor, ON (CD ROM).

Collins, H. & Weinel, M. 2011, 'Transmuted Expertise: How Technical Non-Experts Can Assess Experts and Expertise'. *Argumentation* 25-3, 401-413.

Daukas, N. 2006, 'Epistemic Trust and Social Location'. *Episteme: A Journal of Social Epistemology* 3.1, 109-124.

Goodwin, J. 2011, 'Accounting for the Appeal to the Authority of Experts'. *Argumentation* 25-3, 285-296.

Jacobs, S. 2005, 'Rhetoric and Dialectic from the Standpoint of Normative Pragmatics', *Argumentation* 14, 261-286.

Hamblin, C. 1970, *Fallacies.* London: Meulen.

Olmos, P. 2008, 'Situated practices of testimony. A rhetorical approach'. *Theoria* 61, 57-68.

Sperber, D. & D. Wilson 1986/1995, *Relevance: Communication and Cognition*. 2nd edition. Oxford: Blackwell.

Sperber, D. (In press). 'The guru effect', *Review of Philosophy and Psychology*. Online at: http://www.sperber.com

Sperber, D. Clément, F., Heintz, CH., Mascaro, O., Mercier, H. Origgi, G. & Wilson, D. 2010, Epistemic Vigilance. *Mind & Language,* Vol. 25, No. 4, 359-393

Tindale, Ch.W. 2011, 'Character and Knowledge: Learning from the Speech of Experts'. *Argumentation* 25-3, 341-353

Walton, D. 1997, *Appeal to Expert Opinion: Arguments from Authority.* Pennsylvania: Pennsylvania State University Park.

Walton, D., Reed, Ch. & Macagno, F. 2008, *Argumentation Schemes*. Cambridge: Cambridge University Press.

Willard, C.A. 1990, 'Authority'. *Informal Logic* 12, 11-22.

Zenker, F. 2011 'Experts and Bias: When is the Interest-Based Objection to Expert Argumentation Sound?' *Argumentation* 25-3, 355-370

TEORÍA DE LA ARGUMENTACIÓN, PRINCIPIOS METAFILOSÓFICOS Y CONTROVERSIAS EN FILOSOFÍA

XAVIER DE DONATO
Depto. de Loxica e Filosofía Moral. Univ. de Santiago de Compostela.
Grupo Episteme.
xavier_donato@yahoo.com

ABSTRACT: In the present paper, we discuss some issues related to an old problem: the nature of philosophical controversies and the possibility of progress in philosophy. We discuss some philsophical positions in face of philosophical disputes and some metaphilosophical problems regarding the way in which they are treated by different philosophical schools. We try to avoid incommensurability between different philosophical paradigmas by resorting to common principles of argumentation based in recent argumentation theories.
KEYWORDS: argumentation, philosophical controversies, disagreement, metaphilosophical principles, Rescher.

Como ya advirtió Descartes, la "diferencia de opiniones" en filosofía es una situación perfectamente normal y el consenso en torno a un mismo tema, difícil:

> "Je ne dirai rien de la philosophie, sinon que, voyant qu'elle a été cultivée par les plus excellents esprits qui aient vécu depuis plusieurs siècles, et que néanmoins il ne s'y trouve encore aucune chose dont on ne dispute, et par conséquent qui ne soit douteuse, je n'avois point assez de présomption pour espérer d'y rencontrer mieux que les autres; et que, considérant combien il peut y avoir de diverses opinions touchant une même matière, qui soient soutenues par des gens doctes, sans qu'il y en puisse avoir jamais plus d'une seule qui soit vraie, je réputois presque pour faux tout ce qui n'étoit que vraisemblable." (Descartes 1637/1996, 8).

Esta situación, aunque normal, le parecía, sin embargo, no deseable y, de ahí, que Descartes se viera conducido, según su propia confesión, patente en el texto, a buscar un camino sólido (el método cartesiano) para llegar a una base igualmente sólida en la que fundar la arquitectura del conocimiento. También David Hume nota la ubicuidad del desacuerdo y la controversia en filosofía y, parecidamente, es esta situación la que le lleva a elaborar la filosofía crítica que luego inspiraría a Kant su despertar del sueño dogmático:

> "The most trivial question escapes not our controversy, and in the most momentous we are not able to give any certain decision. Disputes are multiplied, as if every thing was uncertain; and these disputes are managed with the greatest warmth, as if every thing was certain" (Hume 1739/1967, XVIII).

El escepticismo cartesiano era metódico, una suerte de *epojé* con el fin de levantar mejor las bases de su edificio filosófico. En el caso de Hume, su escepticismo surgía del convencimiento de que la escasa solidez de la mayoría de las opiniones previas sugería la necesidad de realizar una profunda crítica del conocimiento y de la filosofía misma.

Esta preocupación, casi inherente a la filosofía misma desde sus orígenes en Grecia, se mantiene hoy día en vista no sólo de la discrepancia entre las diversas tesis y "soluciones" a los problemas filosóficos, sino en vista a la sensación de *dejà-vu* que uno tiene ante el ir y venir de las diferentes tesis filosóficas que surgen, resurgen y se reproducen en formas diversas aún después de haber sido supuestamente "refutadas". Esto nos lleva a preguntarnos por el estatuto de la filosofía y de la argumentación filosófica: ¿es posible un progreso en filosofía? ¿Es posible una argumentación segura y cogente capaz de conseguir asentimiento de todas las partes en disputa independientemente de los principios filosóficos que estas defienden? ¿Es la filosofía, por lo menos en este punto, diferente de la ciencia o, por el contrario, la posibilidad de refutación y de consenso entre las partes es algo tan propio de la ciencia como de la filosofía?

En términos generales y según la variante pirrónica del escepticismo, no podemos creer (ni descreer) ninguna tesis (y, por tanto, debemos suspender el juicio), ya que en principio siempre sería posible diseñar un argumento cogente a favor de la tesis contraria. Aunque hay muchas razones para no ser un escéptico pirrónico con respecto al conocimiento ordinario, podría parecer que la historia de la filosofía, con las ideas y venidas de las mismas tesis y de sus opuestas, nos provee de un ejemplo con respecto al cual acaso sería posible mantener un escepticismo tal. Una forma de negar el pirronismo con respecto a la filosofía sería mantener que en la filosofía puede haber un cierto progreso (al menos para ciertas cuestiones) en el sentido de que las controversias filosóficas pueden resolverse (o disolverse) de un modo objetivo. Esto nos lleva a plantear la cuestión metafilosófica de cómo encarar las controversias filosóficas, tema harto intrincado si tenemos en cuenta que, como dijo en alguna ocasión Albert Camus, lo que para algunos es un axioma indudable para otros es un excelente ejemplo de petición de principio. Igualmente, en los debates filosóficos hacemos uso constante de hombres de paja y cargas de la prueba echadas sobre las espaldas del contrario, sin que a veces parezca haber un juez neutral que dirima la cuestión. La cosa se complica cuando los filósofos que ponemos frente a frente pertenecen a escuelas muy distintas, de modo que ni siquiera comparten los mismos métodos ni objetivos (estaríamos ante casos de inconmensurabilidad radical, parecidos a aquellos que Kuhn y Feyerabend creyeron ver en la historia de la ciencia). Desde la metafilosofía y en un marco pluralista, Nicholas Rescher (1985) y (2006) ha dado claves para afrontar esta cuestión que nos permiten ver cómo podríamos resolver algunas aporías reconociendo, al mismo tiempo, la inevitabilidad de algunos conflictos (aporías).

Precisamente, autores como Rotenstreich (1980) han visto en el carácter inherentemente parcial de los sistemas filosóficos la explicación de la falta de consenso y la propia necesidad de la controversia en filosofía (sin llegar a ver en esto una razón suficiente para sostener una tesis de tipo relativista).

Sin embargo, los principios metafilosóficos de Rescher (2006), extraídos de la larga tradición filosófica y que supuestamente han gobernado la racionalidad de las controversias filosóficas desde los orígenes, sólo sirven como guías o ideales regulativos a los que la buena intención de los filósofos (el "bon sens" del que hablaba Pierre Duhem cuando se trataba de encontrar el consenso entre los científicos) puede recurrir en el deseo de encontrar un camino hacia el asentimiento y la solución de los conflictos. No obstante, estos principios, tales como el "non explicari obscuris per obscurior" o el "non sunt multiplicanda entia sine necessitate" no son susceptibles de una interpretación unívoca. Necesitamos principios pragmáticos, esto es, dependientes del contexto, para poder guiar mejor la discusión filosófica de un modo que resulte igualmente justo con las distintas partes y no sea susceptible de interpretaciones interesadas. Ahora bien, ¿qué es lo que puede ayudar a resolver discusiones filosóficas?

Recurrir a hechos es una posibilidad mientras la discusión filosófica sea en torno a algo respecto de lo cual la experiencia puede darnos algún tipo de información relevante. Aun si la cuestión es meramente conceptual, y en la medida en que uno está de acuerdo con la metodología de la llamada "filosofía experimental", el recurrir a los "hechos" puede ser no sólo útil, sino deseable en ciertos contextos. Pero normalmente, no son los "hechos" los que ayudan a decidir cuestiones filosóficas. La lógica formal es decisiva en los aspectos de corrección formal, pero poco tiene que decir cuando se trata de evaluar la plausibilidad de las premisas y de la conclusión y aún el alcance probatorio de la inferencia (cuando se trata de una inferencia no deductiva). Es así como muchos autores (cf. Johnstone 1954) han visto en la argumentación persuasiva y la retórica el camino adecuado para fundamentar el entendimiento entre los distintos sistemas filosóficos.

Tampoco parece ser ésta una auténtica vía para el entendimiento filosófico, pues la persuasión no parece un medio objetivo y neutral (depende de los sujetos involucrados en una discusión). Hay que recurrir, pues, a algo distinto de los principios metafilosóficos y la retórica, y ese algo sólo puede ser, en mi opinión, una teoría de la argumentación (i) normativa en el sentido de dar normas adecuadas para poder guiar la discusión racional, (ii) pragmática en el sentido de hacer esas normas dependientes del contexto, y (iii) dialógica, en el sentido de dar cabida a las distintas partes de una discusión y concebir sus diferentes tesis o afirmaciones como aportaciones a una discusión en la que intervienen distintos argumentadores, que piden y aducen razones unos a otros. Es por esta razón que creo interesante recurrir a teorías como las de van Eemeren y Grootendorst

(2004) o van Eemeren (2010), con el fin de aunar las dimensiones dialécticas y retóricas de la argumentación en el contexto de los debates filosóficos. No sólo eso, creo que habría recurrir a un análisis filosófico de tipo epistemológico acerca de cuestiones como la carga de la prueba o el uso de los experimentos mentales, por ejemplo, ya que son elementos habituales en las discusiones filosóficas (y también las no filosóficas) y, sin embargo, no encuentran su lugar apropiado en las teorías de la argumentación disponibles hoy día. Así, pues, frente a sugerencias como la de Laudan (2006), quien ve en la filosofía del derecho un ámbito para suministrar modos de dirimir disputas en filosofía o para determinar quién tiene la carga de la prueba en una discusión, nuestra propuesta es acudir a la teoría de la argumentación y al propio análisis conceptual y epistemológico para determinar la cogencia de las argumentaciones filosóficas y la plausibilidad y razonabilidad de las premisas en las que se basan.

Agradecimientos:

Este trabajo ha sido parcialmente financiado por el Ministerio de Economía y Competitividad del gobierno de España (Proyecto de Investigación Ref. FFI2009-08828/FISO).

REFERENCIAS

Descartes, R. 1937/1996, *Discours de la méthode*, en *Ouvres de Descartes*, ed. Adam-Tannery, vol. VI, Vrin, Paris.

Hume, D. 1739/1967, *A Treatise of Human Nature*, ed. L.A. Selby-Bigge, Clarendon Press, Oxford.

Johnstone, H.W. 1954, "The Nature of Philosophical Controversy", *The Journal of Philosophy*, 51, 10, 294-300.

Laudan, L. 2006, *Truth, Error and Criminal Law*, Cambridge University Press, Cambridge.

Rescher, N. 1985, *The Strife of Systems*, Univ. of Pittsburgh, Pittsburgh

Rescher, N. 2006, *Philosophical Dialectics*, State University of New York Press, Albany.

Rotenstreich, N. 1980, "Relativity and Variety of Philosophical Systems", *Philosophy and Phenomenological Research* 41, 187-203.

van Eemeren, F.H. 2004, *A systematic theory of argumentation*, Cambridge University Press, Cambridge.

van Eemeren, F.H. 2010, *Strategic Maneuvering in Argumentative Discourse*, John Benjamins Publ. Co., Amsterdam.

EL LUGAR DE LA ARGUMENTACIÓN EN LOS ESTUDIOS DE FILOSOFÍA. [2]

HUBERT MARRAUD
Universidad Autónoma de Madrid.
hubert.marraud@uam.es

ABSTRACT: My contribution will address the following four questions:
1. What attention is paid to the practice of argumentation in philosophical training?
2. What attention should be paid to the practice of argumentation in philosophical training?
3. What attention is paid to the theory of argumentation in the practice of philosophy?
4. What attention should be paid to the theory of argumentation in the practice of philosophy?
KEYWORDS: argumentation, higher education, logic, philosophy.

El título de esta comunicación podría reformularse como una pregunta: ¿cuál es el lugar de la argumentación en los estudios de filosofía? La pregunta puede servir para plantear debates diferentes, aunque conexos, acerca de las relaciones entre argumentación y filosofía. La pregunta puede entenderse como una cuestión de hecho o como una cuestión de derecho; además debemos distinguir entre la práctica y la teoría de la argumentación.

-¿Qué atención se presta a la práctica de la argumentación en los estudios de filosofía?

-¿Qué atención debe prestarse a la práctica de la argumentación en los estudios de filosofía?

-¿Qué atención se presta a la teoría de la argumentación en los estudios de filosofía?

-¿Qué atención debe prestarse a la teoría de la argumentación en los estudios de filosofía?

Estructuraré mi comunicación al hilo de estas cuatro preguntas.

¿Qué atención se presta a la práctica de la argumentación en los estudios de filosofía?

Para abordar esta cuestión voy a examinar las competencias del libro blanco del título de grado en Filosofía de la ANECA. El libro blanco refleja cierto consenso en la comunidad filosófica española, porque en su elaboración participaron representantes de 27 universidades y porque ha servido como base para el diseño de los grados en Filosofía de las distintas universidades.

[2] Trabajo realizado en el marco del proyecto financiado por la Secretaría de Estado de Investigación, Desarrollo e Innovación (MEC), FFI2011-23125.

El libro blanco divide las competencias en generales y específicas, y éstas en competencias de primer y segundo ciclo, valorándolas de 0 a 4. En las tablas 1 y 2 se recogen las 10 competencias más valoradas de cada grupo y su puntuación. He destacado en negrita las competencias que me parecen intrínsecamente ligadas a la práctica argumentativa.

COMPETENCIAS GENERALES	
Comunicación propia lengua	3,97
Capacidad de reflexión	3,94
Capacidad de enseñar	3,94
Razonamiento crítico	**3,94**
Capacidad de análisis y síntesis	3,74
Capacidad de deliberación	**3,74**
Análisis lógico	**3,73**
Conocimiento y manejo lengua extranjera	3,71
Aprendizaje autónomo	3,65
Habilidades de investigación	3,65

Tabla 1. Competencias generales en el libro blanco

COMPETENCIAS ESPECÍFICAS			
PRIMER CICLO		SEGUNDO CICLO	
Utilizar las bibliotecas con eficacia	3,68	Pensamiento crítico e independiente	3,97
Relacionar ideas	3,58	Examinar problemas	3,93
Plantear preguntas	3,55	Relacionar preguntas	3,93
Sensibilidad a la diversidad de opiniones, prácticas y modos de vida	3,47	**Presentar una evaluación de consideraciones relevantes**	3,90
Sensibilidad para interpretar textos de diversas épocas o tradiciones	3.40	Enfrentarse críticamente a las ideas	3,90
Escuchar presentaciones o exposiciones complejas	3,53	Plantear preguntas	3,90
Examinar problemas	3,33	**Evaluar con claridad y rigor los argumentos de un texto**	3,90
Construir argumentos	3,32	Criticar y reinterpretar	3,87
Sintetizar, analizar y construir argumentos correctos...	3,29	**Construir argumentos**	3,87
Evaluar con claridad y rigor los argumentos de un texto	3,29	Utilizar las bibliotecas con eficacia	3,87

Presentar una evaluación de consideraciones relevantes	3,29		

Tabla 2. Competencias específicas en el libro blanco

De estas tablas se desprende que la formación en filosofía es, en una proporción significativa, adiestramiento en argumentación. En ese adiestramiento intervienen competencias relacionadas con el uso de argumentos y con el análisis y evaluación de argumentos.

Un dato complementario es si la breve presentación del grado que puede encontrarse en la página web de cada universidad menciona la argumentación. La consulta de las páginas web de las 11 universidades con mayor matrícula en la licenciatura en Filosofía el curso 2003-2004 arroja resultados menos concluyentes.

-No lo mencionan la Complutense, Baleares, Salamanca, Sevilla y Valencia.

-La Universidad Autónoma de Barcelona declara que "el plan de estudios incorpora las competencias relacionadas con… el descubrimiento de los presupuestos tácitos de todo discurso, la construcción de argumentos convincentes y la argumentación clara y persuasiva".

-La Autónoma de Madrid indica que el grado en Filosofía persigue, entre otras cosas, una formación básica en "lógica y argumentación" y "adquisición de aptitudes de deliberación".

-La Universidad de Barcelona incluye entre los objetivos del grado "Proponer… la construcción y la evaluación de argumentos".

-La Universidad de Granada señala que "desde los orígenes de la civilización occidental y de las instituciones académicas, … la Filosofía se ha ocupado de los problemas universales de la existencia humana mediante el ejercicio del pensamiento libre, la argumentación racional y el debate público"

-Finalmente, el estudiante de la UNED "ha de mostrar capacidad de diálogo y comprensión de las diversas argumentaciones, de análisis y reflexión" y estar preparado "para poder argumentar y atender argumentos".

Como conclusión, diría que cuando se describe la filosofía como un saber hacer las competencias relacionadas con la argumentación ocupan una posición central, pero que la concepción predominante en la Universidad española entiende la filosofía como un saber ligado a contenidos específicos. Si es así, cabe esperar que en los planes de estudio haya pocas asignaturas dedicadas específicamente a enseñar a argumentar.

¿Qué atención se presta a la teoría de la argumentación en los estudios de filosofía?

A juzgar por los nombres de las asignaturas, hay contenidos de argumentación en aproximadamente la mitad de los grados en Filosofía. No obstante, es engañoso, porque con nombres como "Lógica y teoría de la argumentación" pueden ofrecerse cursos estándar de lógica formal o de historia de la lógica.

Materia	Universidades
Argumentación y retórica	1
Filosofía y argumentación	1
Lógica y argumentación	2
Lógica y teoría de la argumentación	4
Metodología y argumentación lógica	1
Pensamiento crítico	1
Retórica y teoría de la argumentación	1
Teoría de la argumentación	2
Teoría y práctica de la argumentación	1

Tabla 3. Denominaciones de las asignaturas

El reconocimiento del valor instrumental de la argumentación para el aprendizaje y ejercicio de la filosofía solo justifica dar cabida a la teoría de la argumentación en los estudios de filosofía si se cree que conocerla mejora las aptitudes argumentativas de la gente, algo que no es obvio.

El reconocimiento valor instrumental de la argumentación se refleja en dos tipos de asignaturas. En primer lugar, asignaturas de pensamiento crítico, orientación que Luis Vega (2003, p.27) define como "un conjunto de habilidades discursivas básicas que permitan al estudiante afrontar desde el análisis de discusiones en el Congreso hasta el ejercicio reflexivo de la argumentación en clase o la composición de disertaciones por escrito. Pero más allá de unos elementos lingüístico pragmáticos y lógico-informales, no se le conocen mayores pretensiones conceptuales o aspiraciones teóricas". Los planes de estudio de la Universidad de Valencia (*Pensamiento crítico*) y la UAB (*Teoría y práctica de la argumentación*) incluyen asignaturas que, en mayor o menor medida, responden a este planteamiento. Por cierto que un buen indicador de esa orientación es la mención en la bibliografía básica del libro de Weston *Las claves de la argumentación*.

Una alternativa al pensamiento crítico es el aprendizaje a partir de ejemplos, sustentado en la idea de que el mejor modo de que el estudiante de filosofía adquiera las destrezas argumentativas que necesita es el estudio y análisis de argumentos filosóficos paradigmáticos. La Universidad de Granada, en cuyo grado en Filosofía se imparte una asignatura básica llamada *Filosofía y argumentación*, ilustra este planteamiento.

Una razón para estudiar la teoría de la argumentación en el grado que no depende de su carácter propedéutico es considerar que forma parte de la filosofía, como hacen Bermejo Luque (2003) o Hitchcock (2000, p.129): "la lógica informal es una nueva subdisciplina de la filosofía, que puede definirse de forma aproximada como filosofía de la argumentación". La filosofía de la argumentación sería equiparable a este respecto a la filosofía de la técnica o la filosofía de la acción. Creo que las asignaturas que responden más claramente a esta idea son la *Lógica y teoría de la argumentación* de la Universidad de Valencia y la *Teoría de la Argumentación* de la UNED. Es significativo que ninguna de estas asignaturas, a diferencia de las mencionadas antes, sea básica (son obligatoria y optativa, respectivamente).

¿Qué atención se presta a la teoría de la argumentación en los estudios de filosofía?

Simplificando, podemos encontrar tres tipos de asignaturas de argumentación en los planes de estudio del grado en Filosofía: asignaturas de pensamiento crítico, asignaturas de argumentación filosófica y asignaturas de filosofía de la argumentación. Se trata de tipos ideales, en la práctica normalmente nos encontraremos con un *coupage* que combina en distintas proporciones esas tres variedades. Para complicarlo más, la lógica formal se presenta a veces como una teoría de la argumentación o se justifica dentro del *curriculum* filosófico apelando a competencias argumentativas, aportando así una nueva variedad. *Lógica y teoría de la argumentación* es un curso de filosofía de la argumentación en Valencia, un curso de lógica formal en Oviedo, y un curso de historia de la lógica en La Laguna.

El cuadro siguiente da una idea de la situación actual en las 22 universidades españolas, públicas y privadas, que imparten grados en Filosofía.

Materia	Universidades	%
Argumentación filosófica	1	4,5
Lógica formal	21	95,4
Lógica y teoría de la argumentación	3	13,6
Pensamiento crítico	1	4,5
Teoría de la argumentación	7	31,8

Tabla 4. Según los programas de las asignaturas

Se han contabilizado únicamente asignaturas básicas y obligatorias. A estas hay que añadir la presencia de asignaturas optativas en los grados en Filosofía de la UNED (*Teoría de la argumentación*) y la Universidad de Valladolid (*Temas de lenguaje y argumentación*). En suma, hay contenidos obligatorios de argumentación en aproximadamente un tercio de las universidades que ofrecen un grado en Filosofía.

¿Qué atención debe prestarse a la teoría de la argumentación en los estudios de filosofía?

Evidentemente la respuesta depende del modo en que se conciban la filosofía y la teoría de la argumentación. Para responder abordaré dos cuestiones relacionadas: ¿Es la filosofía una práctica argumentativa? ¿Es la teoría de la argumentación filosófica?

¿Es la filosofía una práctica distintivamente argumentativa?

Los debates sobre las relaciones entre argumentación y filosofía suelen girar en torno a la cuestión de si la filosofía es una actividad distintivamente argumentativa. La argumentación está presente en cualquier práctica comunicativa que persiga fines cuya consecución requiera el acuerdo de los participantes, y por consiguiente también en la filosofía. Pero lo que se pregunta es si la filosofía tiene alguna relación privilegiada con el razonamiento y la argumentación, que no tengan la historia, la economía o la química.

Luis Vega (2006) distingue cuatro posiciones sobre la relación entre filosofía y argumentación.

(1) La hipótesis nula, que niega que la argumentación tenga una significación especial para la filosofía. Las distintas orientaciones filosóficas tienen su propio su aparato conceptual, sin que exista una *koiné* a la que traducir las distintas teorías. Como no puede haber un fondo común de acuerdos sobre supuestos y procedimientos, el diálogo entre las distintas escuelas es imposible y la argumentación no desempeña ningún papel relevante.

(2) La hipótesis mínima considera que la argumentación es un recurso típico del discurso filosófico. Si la argumentación es un recurso típico del discurso filosófico, puede preguntarse si hay argumentos filosóficos propios o típicos. Se ha mantenido que el regreso *ad infinitum*, la reducción al absurdo, los argumentos transcendentales o los experimentos mentales son patrones argumentativos típicamente filosóficos.

(3) Para la hipótesis máxima la argumentación es el recurso definitorio del discurso filosófico. La hipótesis máxima es una generalización a partir de la presunta existencia de argumentos filosóficos

propios y exclusivos: la identificación de ciertos discursos argumentativos como inequívocamente filosóficos determina la identificación del discurso filosófico como inequívocamente argumentativo. (4) Para la hipótesis fuerte la argumentación es un recurso necesario del discurso filosófico. Para Vega la argumentación es un recurso necesario de la práctica de la filosofía en la medida en que la filosofía pretenda ser una empresa intelectual (a) susceptible de evaluación y de aprendizaje, (b) cultivada a través de determinadas tradiciones de pensamiento, y (c) con el propósito de contribuir a la lucidez en asuntos públicos o al desarrollo del conocimiento público. La tesis de Vega no es descriptiva, aunque pueda justificarse en parte recurriendo al devenir histórico de la filosofía o al *consensus gentium*. Vega defiende que se *debe* argumentar en filosofía y lo sustenta en la correspondencia entre las dimensiones teórica y práctica de la argumentación y de la filosofía. La argumentación tiene una doble finalidad de coordinación entre proposiciones y entre proposiciones y acciones. Un argumento es un intento de justificar una proposición y de inducir una actitud o una disposición en el destinatario. Vega establece una "complicidad estrecha" entre estas dos dimensiones de la argumentación y las dos dimensiones básicas de la filosofía, teórica y práctica. En su dimensión teórica la filosofía es una empresa cognitiva, de racionalización interna de ideas y creencias, y en su dimensión práctica una empresa directiva de racionalización de la conducta.

¿Es la teoría de la argumentación una teoría distintivamente filosófica?

Para determinar cuál es el lugar que le corresponde a la teoría de la argumentación en la filosofía no solo hay que establecer si la filosofía es argumentativa, sino también si el estudio de la argumentación es propio de la filosofía de alguna manera y a algún respecto. También aquí hay división de opiniones. Para algunos la noción de argumento solo puede explicarse satisfactoriamente dentro de un tratamiento filosófico de la racionalidad. Para otros el estudio la argumentación no es más filosófico que la lógica formal. Desde este punto de vista, la filosofía de la argumentación no sería una teoría de la argumentación sino una reflexión sobre ella.

El objetivo tradicional de la lógica es la búsqueda de estándares y criterios de bondad argumental. La lógica nació como una rama de la filosofía aunque su matematización desde finales del siglo XIX le fue confiriendo una autonomía creciente (al tiempo que le alejaba de una teoría de los argumentos). Con el paso del tiempo el término "lógica" parece haber diversificado sus sentidos hasta resultar equívoco. Entendida como una teoría de la argumentación, la perspectiva lógica se diferencia por centrarse en los argumentos como productos de la argumentación y por su carácter normativo. La vigencia de estos dos rasgos es desigual. Las teorías contemporáneas de la argumentación se caracterizan por un triple giro pragmático (de los argumentos al uso de los argumentos), dialéctico (de las reglas de inferencia a las reglas de procedimiento) y epistémico (de

la inferencia a la justificación). Aunque pasar del estudio y clasificación de los argumentos al de las argumentaciones puede ser novedoso, la normatividad establece una continuidad entre la tradición lógica y corrientes contemporáneas como la lógica informal y la pragmadialéctica. Por tanto, la teoría de la argumentación es filosófica en la medida en que es normativa, lo que dicho sea de paso la asemeja a la ética.

Además la práctica de la argumentación es inherentemente normativa. La forma comunicativa propia de la argumentación es el diálogo: una interacción verbal regulada en la que los participantes tratan de alcanzar un fin compartido intercambiando razones. Las reglas de ese intercambio imponen obligaciones dialécticas a los participantes, como dar razón de sus aserciones si se les pide o responder a las objeciones y contraargumentos. Pero el argumentador no es únicamente responsable de sus aserciones, sino también de la solidez de sus inferencias, y también debe defenderla con razones si se le requiere. Eso diferencia la práctica de la argumentación de la de la natación, puesto que ser capaz de justificar verbalmente nuestra manera de nadar no es nadar. En este sentido alcanzar un determinado nivel de competencia argumentativa exige ser capaz de metaargumentar. Por consiguiente, el conocimiento de una teoría normativa de la argumentación incide directamente en las habilidades argumentativas.

Los metaargumentos tienen una importancia doble para la teoría de la argumentación. Son una clase de argumentos y por tanto objeto de estudio de esa teoría, que si es normativa, consiste esencialmente en metaargumentar: “la evaluación de argumentos solo puede hacerse seriamente si se dan razones que sustenten la tesis valorativa; semejante evaluación razonada es obviamente un argumento, y como su objeto es el argumento original, la evaluación es claramente un metaargumento” (Finocchiaro, 2007, p. 254).

Si estoy en lo cierto hay una continuidad notable entre la teoría normativa y la práctica de la argumentación. Esa continuidad tiene consecuencias importantes:

-el intercambio argumentativo no solo puede cambiar nuestras creencias acerca del mundo, sino también nuestro repertorio de patrones argumentativos y nuestras creencias acerca de cuáles de ellos sean utilizables;

-una teoría normativa de la argumentación no solo evalúa nuestras prácticas argumentativas, sino que también puede cambiarlas.

Los metaargumentos también son importantes en filosofía. Me parece que la argumentación filosófica se caracteriza por la frecuencia con la que es metaargumentativa, con la que se cuestionan patrones argumentativos y garantías inferenciales. Esta conjetura cuenta con avales en la historia de la filosofía: muchos debates escépticos pueden verse como discusiones acerca de la carga de la

prueba y de la comparación de la fuerza de los argumentos, la discusión del argumento del diseño en los *Diálogos sobre la religión natural* de Hume puede verse como una argumentación de segundo orden sobre la fuerza o la debilidad de los argumentos por analogía, etc. También puede aducirse que la capacidad de la filosofía de reflexionar sobre sí misma, sobre sus límites y procedimientos, hace plausible que la argumentación filosófica se caracterice por la frecuencia con la que se vuelve metaargumentativa.

REFERENCIAS

Arroyo, G. 2010, "*Reductio ad Absurdum* en la argumentación filosófica", *Revista Iberoamericana de Argumentación* 1, 1-23.
Bermejo Luque, L. 2003, "Filosofía y retórica: el lugar de la argumentación,. Δαίμων 30, 105-115.
Finocchiaro, M.A. 2005, *Arguments about Arguments,* Nueva York, Cambridge University Press.
-2007, "Arguments, Meta-arguments and Metadialogues: A Reconstruction of Krabbe, Govier and Woods", *Argumentation* 21, 253-268.
Hitchcock, D. 2000, "The Significance of Informal Logic for Philosophy". *Informal Logic* 20, 129-138.
Marraud, H. 2010, "Argumentos suposicionales, razones y premisas", Tópicos 39, 153-165.
Rescher, N. 2001, *Philosophical reasoning. A study in the methodology of philosophizing*, Malden (Mass.) y Oxford, Blackwell.
Vega, L. 2003, *Si de argumentar se trata,* Barcelona, Montesinos.
-2006, "Argumentación y filosofía", *Revista Lindaraja* 6. www.realidadyficcion.eu.

SEIS USOS DE LA ARGUMENTACIÓN.

RAYMUNDO MORADO
UNAM (México)
morado@unam.mx

ABSTRACT: This paper explores six models of argumentation uses, both in the sense of reconstructions of argumentative practices and as recommended ways to go about arguing. We conclude that no specific use, beyond argumentative support, is always present; each one must be invoked in the model as a useful, but inessential use of argumentation.

KEYWORDS: Argumentation, Models, Polemics, Debate, Epideictic argument

Introducción

La teoría de la argumentación, como otras partes de la lógica, puede tener tanto aspectos normativos como descriptivos. Por ello, hay dos preguntas naturales sobre cualquier candidato a modelo de la argumentación: (1) ¿Es un buen modelo a seguir? Es decir, ¿es deseable promover esta manera de argumentar? (2) ¿Es un buen modelo a usar? Es decir, ¿es una buena explicación de lo que ocurre cuando alguien argumenta?

Al inicio de la *Retórica*,[3] Aristóteles reconoce que podemos tratar de convencer epidícticamente (alabando o censurando), y que argumentamos para exhortar o disuadir.
Sin embargo, considera que la argumentación forense o jurídica solamente puede ser "o acusatoria o defensiva, pues los litigantes deben necesariamente acusar o defender".[4] Así pues, desde el inicio de los estudios sobre argumentación, es claro por un lado que hay varios tipos de discurso retórico argumental. Por otro lado, es natural concentrarnos en el estilo polémico y utilizarlo para modelar a toda la argumentación.

Necesitamos modelos de la argumentación que hagan justicia a las diferentes funciones, propósitos y dinámicas argumentativos. El peligro es un modelo que considere nada más un uso. Aunque tal uso pueda ser valioso para ciertos tipos de discurso puede ser insuficiente para modelar la diversidad de las prácticas argumentativas. Pasaremos revista a diferentes usos de argumentación

[3] 1358b, por ejemplo.
[4] Íbid.,
δίκης δὲ τὸμὲν κατηγορία, τὸ δ᾽ ἀπολογία: τούτων γὰρ ὁποτερονοῦνποιεῖν ἀνάγκη τοὺς ἀμφισβητοῦντας ".

explorando sus riquezas y limitaciones para concluir en la necesidad de modelos incluyentes y flexibles.

1. La argumentación como disputa:

Hay la opinión de que argumentar es defender una posición como en la guerra. Esta idea bélica de la argumentación es muy común en nuestra sociedad incluyendo los ámbitos académicos. Se considera muy útil tanto para usarla, como para seguirla. Según muchos profesores, los alumnos deben luchar por sus ideas pues ese modelo no solamente explica toda argumentación sino que es la manera justificada de plantear cualquier estrategia argumentativa.

Esta práctica está salpicada de metáforas guerreras. En palabras de Cohen (p. 178), "We routinely speak, for example, of knockdown, or even killer, arguments and powerful counterattacks, of defensible positions and winning strategies, and of weak arguments that are easily shot down while strong ones have a lot of punch and are right on target."

Esas metáforas no deben tomarse a la ligera. Las palabras reflejan el pensamiento que nos dirige. Como insisten Lakoff y Johnson, "The fact that we in part conceptualize arguments in terms of battle systematically influences the shape arguments take".[5]

Seguir este modelo puede ser muy destructivo para la práctica social de la argumentación, llevando a veces a la violencia verbal innecesaria. Afortunadamente, este modelo no se aplica a todos los tipos de argumentación, muchos de los cuales no requieren atacar a otros para defender una posición. Por ejemplo, Cohen (p. 185) trata de liberarnos de las metáforas bélicas en nuestra idea de la argumentación y propone que "Argument is not war; it is diplomatic negotiation."

2. La argumentación como debate:

Un segundo modelo del argumentar es concebirlo en términos de defender una posición, aunque en un sentido no bélico. Se argumenta con un oponente al que se puede estar tratando de ayudar a desembarazarse de ideas erróneas; el otro debatiente no es necesariamente un enemigo, sino tan sólo un rival. Esta idea de la argumentación es muy útil pero hay que seguirla con cuidado pues, también lleva a veces a polarizar y excluir.

Hablar de un oponente, así sea una mera palabra, parece estimular una actitud negativa.

> "When people are adversaries, even when they are adversaries only in virtue of the roles they occupy temporarily, other features very often accompany their opposition: lack of respect, rudeness, lack of empathy, name-calling, animosity, hostility, failure to listen and attend carefully, misinterpretation, inefficiency, dogmatism, intolerance, irritability, quarrelsomeness, and so forth".[6]

[5] "Metaphors We Live By". George Lakoff y Mark Johnson (1980).
[6] Govier (1999), p. 245.

Govier ha criticado con razón al modelo proponente/oponente de la pragmadialéctica por las dificultades que tiene para reconciliar oposición y colaboración en los dialogantes. Las partes resuelven las disputas como colaboradores aunque el modelo de diálogo es como opositores.[7] Ante la respuesta de que "oponente" es solamente una forma de hablar, Govier responde con justificada suspicacia.[8] Si vemos a los participantes más como colaboradores que como opositores, estamos transitando de un modelo de disputa como debate a uno de disputa como controversia.

3. La argumentación como controversia:

Aplicar un modelo de controversia es un progreso respecto a seguir uno bélico o de disputa porque éstos tienden a producir posiciones antagónicas excluyentes. En cambio la controversia admite muchas posiciones que se traslapan y pueden beneficiarse parcialmente de otras aunque sean excluyentes consideradas en su totalidad. Por supuesto, la controversia puede ser externa, con otros participantes, o interna. Se puede asumir posiciones opuestas a las propias creencias para explorar dichas posiciones o, como en el caso de la *Reductio*, para mejor atacarlas.

Esta idea de la argumentación es muy útil pero debe ser usada con cuidado. La controversia sigue teniendo un elemento de oposición, así sea mínimo, pues requiere que haya algún desacuerdo,[9] y éste puede dar lugar a conductas dañinas. "Controversies often involve rudeness, disrespect, hostility, animosity, name-calling, put-down, insults, *ad hominem* attacks, misinterpretation, diversions into unnecessary and irrelevant themes, intolerance, dogmatism, wasted energy, failures of communication, and waste of time and talent. [...] our capacity for action may be inhibited."[10]

El uso de este modelo también es peligroso pues presupone la diversidad de opinión como condición esencial para dar sentido a la argumentación. Para algunos autores, el desacuerdo es un elemento esencial de la noción mínima de argumentación. Willard entiende "By argument I mean a process of dissensual interaction -- conversation involving disagreement."[11] En la pragma-dialéctica se ha dicho que sin antagonista no hay argumento: "Argument [...] is a part of a discourse procedure whereby two or more individuals who have a difference of opinion try to arrive at agreement.

[7] Ver Govier (1999), p. 255, nota 5.
[8] Govier (1999), pp. 201, 255 nota 8 y 256 nota 20.
[9] Govier (1999), p. 244, llama a esto "minimal adversariality".
[10] Govier (1999), p. 248. Hay que notar que Govier, quien señala tantos peligros en la controversia, está propugnando por su uso.
[11] Willard (1987), p. 145.

Argument presupposes two distinguishable participant roles, that of a 'protagonist' of a standpoint and that of a --real or projected-- 'antagonist'."[12]

Pero Govier deja en claro que defender una posición no implica atacar otras: "Although this sliding of categories from difference to disagreement, to conflict and opposition, is natural, it is undesirable and conceptually unnecessary."[13]

Desgraciadamente, Govier no menciona la posibilidad de que estemos haciendo una argumentación no-defensiva. Dice: "One who offers an argument openly acknowledges the actuality or possibility of disagreement or doubt [...] the practice of argument implies a recognition that there may be legitimate doubts".[14] Poco después añade que "A person who offers an argument for a claim acknowledges the existence, or possibility, of dispute".[15]

Govier nota que algunos argumentos que nos damos a nosotros mismos llegan a conclusiones que sostenemos de antemano. Incluso llega a decir que "it is possible to argue in non-confrontational ways".[16] La explicación que encuentra es que usamos los argumentos para investigar cómo justificar alguna tesis importante.[17] Esto no es mucho avance pues quien cree que toda argumentación es defensiva podría preguntar: "¿Para qué buscar cómo justificar una posición, si no es como preparación para su defensa?". Es sintomático que para Govier argumentar nunca puede ser "preaching only to the converted"[18], pues no puede ver que en ese discurso también hay argumentos.

4. La argumentación como fundamentación:

Argumentar a veces es defender una posición como en la guerra, el debate o la controversia. A veces es simplemente defender, sin sentido bélico, sin oponente adversario, y sin colaborador o controversia alguna.

Tal defensa es mejor descrita como una construcción o fundamentación que puede eventualmente servir para colaborar, para defender e incluso para atacar. Seguir este modelo no excluye ninguno de los otros pues se circunscribe al único uso esencial de la argumentación que es el de fundamentar una posición.

12 van Eemeren et al. (1996).
13 *Ibid*., p. 54.
14 *Ibid.,* p. 47.
15 *Ibidem.,* p. 50.
16 *Ibidem.,* p. 55.
17 "as tools of inquiry -- to explore the possibilities of justifying central claims". Govier (1999), p. 46.
18 *Ibidem.,* p. 47.

Los modelos de la argumentación con más elementos como diálogo, controversia, debate, etc., son muy útiles para construir y analizar gran cantidad de discursos e incluso tal vez para todas las instancias normales de argumentación, pero no para todas las instancias de argumentación reales, ni mucho menos para todas las instancias posibles. La diferencia entre las instancias y las instancias posibles es importante porque hay una necesidad de estimular, en ámbitos donde es frecuente la confrontación como el derecho y la filosofía, formas de argumentación no adversarias.

Y debemos estimular en ámbitos colaborativos como la diplomacia y la negociación formas argumentativas imparciales, ni a favor ni en contra de las partes que argumentan. Sea esta imparcialidad psicológicamente posible o no, es metodológicamente un provechoso ideal normativo.[19] Hay que estimular todas las formas dialógicas de la argumentación y también sus formas no dialógicas, cada una de acuerdo con el contexto de uso sin excluir de antemano a ninguna.

Usar este modelo es útil porque no requiere de haber tomado una posición previa. Existe la idea de que una opinión solamente puede ser puesta a prueba exponiéndola a objeciones y críticas. Esta idea es falsa. Muchos argumentos que constituyen demostraciones de opiniones o creencias en geometría o lógica se pueden poner a prueba por un checador automático de teoremas de fácil construcción y completamente carente de opinión adversa a lo que está examinando. Aunque desgraciadamente esto no es posible en toda argumentación, lo es con respecto a suficientes demostraciones sobre temas importantes (al menos para los geómetras y los lógicos) como para reivindicar la noción y la realidad de un tipo de argumentación sin oposición en ningún sentido sustantivo.[20] El punto de fondo es que a veces ni estamos justificando nuestra postura frente a opiniones distintas, ni estamos buscando cómo hacerlo.

Por ello el punto de partida de la argumentación no tiene por qué ser el desacuerdo. Cuando otras personas no comparten nuestras ideas, esto puede tener una gran cantidad de causas. Toulmin, Rieke y Janik identifican tres: porque creen otra cosa, no ven cómo justificar nuestra posición, no tienen bases para tomar una posición.[21] Pero también puede ser el acuerdo. Después de todo, "el desarrollo de la argumentación, así como su punto de partida, implica la aprobación del auditorio."[22]

En el campo de la ciencia, se ha dicho que avanza paradigmáticamente mediante lucha de opuestos (Mill, Popper). Pero también se ha dicho que una teoría científica no triunfa por convencer

[19] Esta idea se la debo a la Dra. Ana María de los Ángeles Ornelas Huitrón.

[20] Con un mínimo de imaginación, siempre es posible alguna metáfora o analogía para extender la noción de "oposición" hasta cubrir casos para los que no usaríamos esta palabra normalmente. Este ejercicio lingüístico es de poco interés teórico.

[21] Toulmin *et al*. (1979), p. 24.

[22] Perelman *et al*., p. 119.

a los contrarios, sino porque los contrarios mueren sin ser reemplazados por nuevos creyentes.[23] Entonces, ¿para quién se está argumentando en ciencia? Puede ser para los colegas que no han tomado todavía posición y que pueden ser ganados a la nueva teoría, no desde la posición contraria sino desde la indecisión. A menudo en ciencia lo que parece una argumentación polémica es en realidad una argumentación apologética o una didáctica, especialmente cuando no hay esperanza de convencer a quienes ya han tomado una posición opuesta.

5.- La argumentación como sistematización

Concepción Martínez Vidal nota que si ampliamos nuestra noción de axioma más allá de las verdades autoevidentes, podemos incluir proposiciones que prometen ser fructíferas, aunque no sean obvias. En este caso, pueden ser los teoremas los que den apoyo a los axiomas: "Los axiomas se aceptan o no por parte de la comunidad matemática en virtud del interés de los teoremas que se pueden probar a partir de ellos"[24] (p. 81). Esta idea de escoger nuestros principios o premisas por su fertilidad no es extraña a la historia de la lógica, por ejemplo la encontramos en las propuestas de definiciones analíticas en Frege.[25]

La lectura de Martínez Vidal me sugiere que a veces el propósito de la argumentación no es defender ni atacar; ni siquiera instruir. Puede ser sistematizar una teoría y explicitar posibles relaciones estructurales entre términos y principios. Por ejemplo, con nuestros argumentos en geometría no estamos discutiendo, estamos explicando. Esto muestra que puede haber explicaciones argumentadas, sin diferencia de opinión. No se trata tan sólo de que la argumentación en la *Ética* de Spinoza pueda ser un monólogo, sino de que su lectura polémica es más un artefacto cultural que una necesidad inherente al texto. Es posible presentar argumentativamente un sistema para realzar el apoyo teórico y los compromisos inferenciales aunque tal sistematización pueda ser un catecismo para los ya convertidos.

6.- La argumentación como discurso epidíctico.

A veces la argumentación aparece en tipos de discurso en que no se está discutiendo; por ejemplo en una boda al alabar o encomiar a los novios; en un funeral al hacer la apología de la persona a quien se despide; al festejar a los triunfadores deportivos o a los egresados que se gradúan y a sus padres, profesores y autoridades. Todos estos tipos de discurso son a veces reunidos bajo el

[23] En palabras de Pascal, "*tous les différends de cette sorte demeurent éternels si quelqu'un ne les interrompt, et qu'ils ne peuvent être achevés si une des deux parties ne commence à finir*". Citado en Granger (1985) p. 343.
[24] En su artículo "Axioma/Axiomatización",
[25] Cfr. Morado (1989)*, passim.*

rubro de discurso "epidíctico". El orador epidíctico "presentaba un discurso al que nadie se oponía, sobre temas que no parecían dudosos y de los que no se sacaba ninguna consecuencia práctica".[26]

Como nota Perelman, el género epidíctico, con todas sus argumentaciones, es preferido por quienes "defienden los valores tradicionales, los valores admitidos, los que constituyen el objeto de la educación, y no los valores revolucionarios, los valores nuevos que suscitan polémicas y controversias. [...] el orador se hace educador."[27] Para Perelman, al educar se argumenta lo no controvertido: "lo que va a decir no suscita controversia [...] no se trata de defender o de atacar".[28]

"Los discursos epidícticos tienen como finalidad aumentar la intensidad de adhesión a los valores comunes del auditorio y del orador."[29] Hablan de "la comunión de los espíritus".[30] Más que de incrementar la intensidad del convencimiento (que puede ya ser total) se trata de armonizar tal convencimiento con las actitudes de la persona. De lo que se trata no es tanto de reforzar un punto de vista como una manera de ver y de apreciar.

"El discurso epidíctico -y cualquier educación- persigue menos un cambio en las creencias que una argumentación de la adhesión a lo que ya está admitido".[31] Mientras que tenemos el modelo del agresor en el debate y el del colaborador en la argumentación inquisitiva, lo que tenemos en el argumento epidíctico es la figura del protector frente a otros sistemas de valores: "se parecerá al guardián protector de los diques que sufren sin cesar el ataque del océano".[32] Pero hay que notar que no se trata de atacar otros valores o defenderse de ellos, sino de confirmar los propios, reforzando los diques.

Perelman describe a la argumentación como un intento "de influir, por medio del discurso, en la intensidad de la adhesión de un auditorio a ciertas tesis".[33] Siguiendo una distinción semejante a la de Perelman[34] podríamos decir que la argumentación puede ser usada para convertir la convicción (creencia en la verdad de una tesis) en persuasión (la adhesión práctica y personal). Le predicamos al coro, ya convencido, pero a menudo todavía no persuadido.

A. C. Baird, distingue la discusión en que se busca heurísticamente la verdad, del debate partidista en que solamente se trata de presentar erísticamente (antagonísticamente) un lado de la cuestión.[35] A esto Perelman comenta que tanto el diálogo heurístico como el erístico son

26 Perelman (1958), p. 95.
27 Perelman (1958), p. 100.
28 Perelman (1958), p. 101.
29 Perelman (1958), p. 102
30 Perelman (1958), p. 105.
31 Perelman (1958), p. 104.
32 Perelman (1958), p. 105.
33 Perelman (1958), p. 48.
34 Perelman (1958), par. 6, pp. 65-71.
35 *Argumentation, Discussion and Debate*, p. 307. Citado en Perelman (1958), p. 81.

excepcionales. Lo normal no es argumentar para buscar la verdad o vencer al adversario sino para motivar a la acción.[36]

Necesitamos buenos argumentos si hemos de presentar nuestras conclusiones públicamente, como es el caso con algunas sentencias judiciales, decisiones políticas y teoremas matemáticos. La argumentación puede racionalizar (en más de un sentido) intuiciones previas. Esta técnica puede devolver a alguien el "orgullo, el equilibrio, la confianza, el consuelo; pone en orden sus ideas, encuadra la decisión y, por consiguiente, la refuerza. Actúa como el teólogo que proporciona las pruebas racionales de un dogma en el que todos los miembros de la Iglesia creían ya con anterioridad".[37]

Conclusiones

Tanto para construir, como para reconstruir los argumentos debemos considerar la diversidad de las funciones de la argumentación. Aplicar el modelo bélico a la argumentación matemática confunde tanto como aplicar el modelo de fundamentación a los debates políticos. Tanto el modelo bélico, el disputante, el de controversia y el de fundamentación tienen su lugar. Ningún modelo se aplica a todos los casos, ninguno deja de iluminar diferentes aspectos de nuestra racionalidad. Ciertamente, es difícil modelar la refutación sin oposición y desde Aristóteles la reconocemos como un tipo indiscutible de argumentación.[38] Estos modelos tradicionales son útiles pero limitados.

Finalmente, podemos ver que la base mínima de la argumentación es abogar por una tesis; pero los aspectos de defensa, ataque, persuasión, diálogo, etc., se deben añadir a medida que se necesiten y resistir todo el tiempo la tentación de tratarlos como elementos esenciales de la modelación.

REFERENCIAS

Aristóteles, *Tratados de Lógica (Órganon) I*. Trad. de Miguel Candel Sanmartín, Madrid, Gredos, 1982.

Cohen, Daniel H. "Argument is War . . . and War is Hell: Philosophy, Education, and Metaphors for Argumentation". *Informal Logic*, Vol. 17, No. 2 (Spring 1995): pp. 177-188.

Gilles Gaston Granger, "Discussing or Convincing: An approach towards a pragmatic study of the languages of science", en Marcelo Dascal (ed.), *Dialogue: An Interdisciplinary Approach*, Amsterdam/Philadelphia: John Benjamins Publishing Company, 1985, pp. 339-351.

[36] Perelman (1958), p. 83.

[37] Perelman (1958), p. 89.

[38] En palabras de Aristóteles, "Es completamente absurdo discutir acerca de la refutación sin hacerlo antes acerca del razonamiento: pues la refutación es un razonamiento", 171a , p. 333. En la traducción de W. A. Pickard-Cambridge, "It is, too, altogether absurd to discuss Refutation without first discussing Proof: for a refutation is a proof". (Part 10, *On Sophistical Refutations*, Aristotle, 350 B.C.E.)

Govier, Trudy. *The Philosophy of Argument*. Virginia: Vale Press, 1999.

Lakoff, George; Mark Johnson. *Metaphors We Live By.* Chicago: University of Chicago Press, 1980.

Perelman, Chaïm; Olbrechts-Tyteca, Lucie (1958), *Tratado de la argumentación*. Cito la versión de Madrid: Gredos, 1994.

Toulmin, Stephen; Rieke, Richard; Janik Allan. *An introduction to reasoning*. New York /London, Macmillan/Collier-Macmillan, 1979.

van Eemeren, Frans H.; Grootendorst y Francisca Snoeck Henkemans (eds.), *Fundamentals of argumentation theory. A handbook of historical backgrounds and contemporary developments*. Hillsdale (NJ), Lawrence Erlbaum Associates, 1996.

Willard, Charles Arthur. "Valuing Dissensus". Capítulo 13 de van Eemeren, Frans H.; Grootendorst, Rob; Blair, J. Anthony; Willard, Charles A. (eds.), *Argumentation: Across the Lines of Discipline*. Dordrecht-Holland/Providence-USA: Foris Publications, pp. 145-157, 1987.

ĒTHOS DIA TOU LOGOU / ĒTHOS DIA TŌN ĒTHŌN: LA EVALUACIÓN DEL CARÁCTER DEL ARGUMENTADOR

PAULA OLMOS GÓMEZ
Dpto. de Lógica, Historia y Filosofía de la Ciencia, UNED.
paulaolm@gmail.com

ABSTRACT: According to Aristotle's *Rhetoric* (1356a9-10), *ēthos*, the orator's character, should be analysed and evaluated through his discourse ("dia tou logou"), not through the previous information and opinions we may have about her acts and person ("alla mē dia to prodedoxasthai"). This caution is, in fact, an Aristotelian technical innovation on the typically Greek widespread rhetorical exploitation of the inconsistencies and contradictions between the speaker's (or, for that matter, the opponent's) words and deeds. The *Rhetoric to Alexander* a contemporary but allegedly more conventional treatise attributed to Anaximenes, is, for example, full of indications to this purpose, in its discussions about enthymemes, *tekmerion* proofs or proofs based on the "opinion about the speaker". The contemporary discussions about complex notions of *ēthos*, either including o excluding aspects as biographical data, public *personae*, author's discursive image etc. in the literary realm might help us clarify some of these issues. Conscious of the complexities of such possible "ontologies of authorship" or "ethotical ontologies", I will claim that although Aristotle's mandate should not be taken as a permanent rule (that would preclude the exposure, in the adequate contexts, of the speaker's imposture), it is nevertheless a useful initial imperative of ethical value, acting as a rebuttable presumption about the speaker's pragmatical consistency that will help us better understand her discourse and attend to certain characteristics about her speaking rights.
KEYWORDS
Anaximenes, argumentation, Aristtle, ethos, persuasion, proof, rhetoric.

Según la *Retórica* de Aristóteles (1356a9-10), el *ēthos* o carácter del orador –una de las tres fuentes básicas de la persuasión, según este mismo texto, junto con el *logos* y el *pathos*– debe entenderse o captarse a partir de los elementos (tanto verbales como no verbales y paraverbales) de su propio discurso ("*dia tou logou*"); y no, se dice explícitamente, de la información u opiniones que previamente tengamos sobre el hablante ("*alla mē dia* <u>*to prodedoxasthai*</u>", literalmente "pero no a partir de <u>lo ya opinado con anterioridad</u>").

Podemos ver, en esta aclaración o "advertencia" aristotélica, el inicio de la larga y compleja tradición retórica del análisis de la auto-representación del hablante/autor, es decir, de la concepción del *ēthos* como "la imagen de sí que el locutor construye, lo quiera o no, <u>a través de su discurso</u>" (Bokobza Kahan 2009, 2; *Cf.* Amossy 1999, 2006 [2000]), algo en principio distinguible de su ser

real o biográfico.[39] Los teóricos contemporáneos se esfuerzan, a partir de tales distinciones, en estudiar las maneras de abordar las relaciones entre dicho *ēthos* y a) su propio texto o discurso, b) su yo biográfico o social y c) el medio social o institucional de su práctica discursiva, en lo que supone una "concepción compleja del *ēthos*".

En el caso de R. Amossy (1999), por ejemplo, se trataría, no tanto de identificar el *ēthos* alternativamente con los polos extremos del yo biográfico o la pura imagen discursiva, de manera excluyente, sino de articular una aproximación integral al concepto como constelación de relaciones que incluya tanto elementos textuales como extratextuales; una concepción que admita que el *ēthos*, que se construye básicamente mediante el discurso, se nutre, sin embargo, de la historia biográfica y de la situación empírica del locutor en el campo social en el que éste actúa, con sus características y posicionamientos políticos, institucionales etc. Es por lo que en trabajos posteriores menciona la "doble naturaleza de la imagen del autor" ("entre le texte et le hors-texte", Amossy 2009, 11), complicando su análisis, por otro lado, con la constatación de las diferencias entre lo construido/proyectado/pretendido por el locutor y lo finalmente atribuido/elaborado/recibido por terceros (también situados en distintas posiciones y roles) que quedaría más o menos fuera de su control: "une multiplicité d'images souvant diverses et contradictoires circulent en proposant du même auteur un kaléidoscope mouvant" (Amossy 2009, 4).

Los estudios literarios son, sin duda lo que más han avanzado (y se han enredado) en la exploración de los distintos "juegos" ethóticos que pueden describirse y analizarse a través de las nociones (por otro lado con variables históricas) de "autor", "imagen de autor", "autor-responsable" (que ha de responder), "autor-actor", "autor-*auctor*", etc.–véase, *e.g.*, Mainguenau 2009, que habla de una "ontología de la autoría" que para nuestros intereses podríamos interpretar como una "ontología del *ēthos*". A partir del influyente trabajo de Wayne C. Booth (1961) sobre el carácter ineludible del autor implícito (*implied author*) en todo texto, dichas ontologías y distinciones se multiplican en los diversos estudiosos del tema. Otro ejemplo significativo sería el de J. Meizoz (2009, 6) quien distingue, en primer lugar, entre a) persona (individuo o ente civil), b) escritor (autor en el campo literario) y c) "inscriptor" o enunciador textual (narrador en un sentido amplio). Lo que le permite, más adelante organizar su propia "ontología de la autoría" conforme a otras tres nociones: i) el *ēthos* discursivo (ethos discursif), que sería la noción más ligada al "inscriptor", por "surgir de un análisis interno del texto", ii) la imagen de autor (image d'auteur) que relacionaría dicho análisis interno con la "información de que el lector dispone sobre dicho escritor", es decir,

[39] Para una revisión actual de los diversos aspectos que comporta, para la teoría literaria y de análisis del discurso, la figura del autor, ver el número monográfico bajo el título de "Ethos discursif et image d'auteur" de la revista *Argumentation et Analyse du Discours* 3 (2009), ed. M. Bokobza Kahan.

pondría en relación el escritor con el inscriptor y iii) la posición o propuesta (posture) del autor como noción compleja que supera la distinción interno/externo y se ocupa de la "conducta del escritor en relación con el discurso del inscriptor y las acciones de su persona (real)". Como podemos ver, *ēthos* tiene un sentido diverso (casi opuesto) en este autor (más cercano a la "advertencia" de Aristóteles) al que presenta Amossy, para quien *ēthos* sería la noción global, casi equivalente a la de "posture" en términos de Meizoz. Lo importante es saber, en todo caso y en cada momento a qué nos referimos y ser consciente de la complejidad de nuestros referentes.

Pero volvamos a Aristóteles y a sus intenciones explícitamente retóricas, es decir, su pretensión de analizar, evaluar y proponer "lo que en cada caso es pertinente para la persuasión" (*to endexomenon pythanon*, *Ret.* 1355b25-26). En lo relativo al "carácter del orador" como fuente persuasiva (no ya como propuesta literaria) su advertencia de que debemos centrarnos en su "*ēthos* discursivo", prescindiendo de las informaciones u opiniones previamente obtenidas o concebidas, no parece en principio estar muy justificado en relación con un auditorio que ha decidir sobre cuestiones civiles (referencia fundamental de la retórica antigua): ¿por qué habríamos de renunciar a dicha información en nuestra recepción o confrontación del discurso? De hecho, por muchas resonancias retóricas que dicha advertencia pueda albergar para el lector moderno (mucho más habituado a los juegos y teorías varias sobre literatura, polifonía, performatividad, teatralidad, máscaras, roles, *personae* etc…), Aristóteles está, de hecho, introduciendo un cambio de perspectiva original y altamente innovador sobre la doctrina griega vigente en el s. IV a.n.e., tal como viene representada por la más o menos contemporánea y supuestamente más convencional[40] *Retórica a Alejandro*, obra atribuida hoy en día casi unánimemente a Anaxímenes de Lámpsaco, y por lo que sabemos de las enseñanzas de Isócrates.[41]

El rendimiento que este texto obtiene, precisamente, de la confrontación y posibles discordancias entre los discursos y las acciones y caracteres "conocidos" (por otros medios) de los interlocutores o de los personajes que forman parte de la controversia (particularmente en el contexto legal) supone, de hecho, uno de los núcleos básicos de la técnica argumentativa que promueve. La *Retórica a Alejandro* enfatiza, en efecto, en varios momentos y a propósito de varios tipos de recursos argumentativos (varios "tipos de prueba") la conveniencia de contar con información sobre el interlocutor o las personas de las que hablamos para destacar y explotar las posibles inconsistencias pragmáticas entre su discurso y sus acciones o costumbres (*ēthoi*).

[40] Chiron (2011) niega el convencionalismo del texto y habla de un clima de innovación al que pertenecen tanto la *Retórica* de Aristóteles como la *RaA* y las propias enseñanzas vivas de Isócrates, en un clima de competitividad local por la supremacía en la enseñanza de la retórica.

[41] Noël. (2011) destaca explícitamente esta oposición en la manera discursiva de entender lo ethótico en Aristóteles (*dia tou logou*) frente a la de Anaxímenes e Isócrates que sería más bien, en expresión paralela *dia tōn ēthōn* o mediante las propias acciones.

Veámoslo a través del reciente estudio de M. Kraus (2011), que nos ofrece la siguiente clasificación de los "tipos de prueba" mencionados, comentados y propuestos en la *Retórica a Alejandro*:

1. Pruebas (*pisteis*) obtenidas por el orador del núcleo de la controversia, i.e., de los discursos, acciones e informaciones sobre personas involucradas.
    a. *eikos* (basada en lo probable o plausible)
    b. *paradeigma* (ejemplo)
    c. *tekmerion* (indicación de una inconsistencia "pragmática")
    d. *enthymema* (entimema o "indicación de una contradicción")
    e. *gnomē* (máxima)
    f. *sēmeion* (signo)
    g. *elenchos* (refutación efectiva directa)

2. Pruebas aportadas o externas (*epithetoi*)
    a. *doxa tou legontos*
    b. testigos
    c. torturas
    d. juramentos

El interés por el tipo de información extra-discursiva de la que hablamos podría aparecer ya en el *enthymema* (en términos de Anaxímenes) aunque lo hace ya de manera muy explícita en el *tekmerion*.[42] Obsérvese, en todo caso, que el significado de estos términos no es equivalente a su uso aristotélico y ni al tradicional posterior. De hecho, Noël (2011) destaca precisamente el importante papel que desempeña en el texto de Aanaxímenes el modelo de contra-argumentación llamado *tekmerion* y que suele oponerse a un argumento rebatible previo (basado en algo plausible *eikos* o en algún signo *sēmeion*). Al señalar una inconsistencia pragmática entre el discurso presente (escuchado en la ocasión retórica determinada) y lo previamente sabido respecto del hablante o de los personajes objeto de su discurso, el *tekmerion* refuta el argumento propuesto y resulta, si no irrebatible, al menos contundente (de dónde le viene el sentido de "signo irrefutable" que obtiene en Aristóteles). El *tekmerion*, también como recurso estilístico de tipo dialéctico (frente al oponente) y retórico (frente al auditorio) mostraría de algún modo "tanto la inconsistencia ética del contrario como la consistencia propia" (Noël 2011: 334). En todo caso estaría explotando una noción del *ēthos dia tōn ēthōn* con referencia al mundo extra discursivo y las características éticas "reales" del hablante.

[42] Es Pisani (2011) el que hace más explícita la diferencia entre *enthymema* y *tekmerion* en la RaA de un modo que indica que también el *enthymema* puede tener que ver con nuestros intereses ethóticos. No es que la contradicción que destaca el *enthymema* sea siempre discursiva (palabras/palabras) y la del *tekmerion* pragmática (palabras/hechos), como podría aparecer a simple vista; sino que el *enthymema* apela más a principios compartidos y destaca "All that departs from commonly accepted ethical and social principles" mientras el *tekmerion* se centra en un personaje concreto y en la posible inconsistencia interna entre *sus* palabras y actos.

Kraus (2011), por su parte, dedica especial atención al tipo de prueba "aportada" (*epithetos*) llamado ambiguamente *doxa tou legontos* (opinión del/sobre el hablante) cuya interpretación ha suscitado debate entre los especialistas.[43] La postura de Kraus (vs. e.g. Mirhady) es que es más plausible que se refiera a la reputación del que habla y no tanto a sus opiniones (2011: 275-278).[44] Ello abre la puerta a algo equivalente al *ēthos* (en su vertiente, sin embargo, más biográfica y opuesta a la "advertencia" de Aristóteles) en la clasificación de la *RaA*, lo que resulta adecuado. Además, Kraus apunta finalmente la posibilidad de que exista una ambigüedad razonable entre el sentido profundo de quién es el que habla y qué viene a defender (algo que más o menos recogería, aunque no muy explícitamente, la traducción por "competencia"). Las dos cosas serían caras de una misma moneda y ambas formarían parte de la *doxa tou legontos*, con lo que esta "fuente de prueba" sería aún más equivalente a un sentido moderno (complejo) del *ēthos* a la vez intra y extra discursivo.

Frente a ello, la clasificación de las pruebas de la *Retórica* de Aristóteles que pretende, en principio, mayor exhaustividad y rigor conceptual nos daría:

1. Pruebas propias del arte (*entechnoi*)
    - a. *logos*
        - i. *enthymema* (basado en lo *eikos* o en *sēmeia*, puede ser irrefutable *tekmerion* o indicativo, sin nombre), incluye como partes las *gnomai*, puede ser refutativo (*elenchos*).
        - ii. *paradeigma* (puede ser ficticio).
    - b. *ēthos*
    - c. *pathos*

2. Pruebas externas al arte (*atechnoi*)
    - a. leyes
    - b. testigos
    - c. torturas
    - d. juramentos
    - e. contratos.

Kraus comenta al respecto la evidencia de un trabajo conceptual aristotélico que supone una "reducción racionalista de tipo filosófico" que hace que todo le encaje al Estagirita de un modo pretendidamente exhaustivo. Habla también de la autoinfluencia de sus obras lógicas, con la dicotomía *syllogismos*/*epagōgē* que se reflejaría en los sentidos por un lado restringidos (en cuanto

[43] Nuestro traductor José Sánchez Sanz ([ed.] 1989) la vierte como "competencia" (expertise) del orador.

[44] Mirhady, según Kraus, tiene una propuesta compleja y también interesante. Dice que la *doxa tou legontos* se corresponde con ciertas opiniones propias del hablante pero no cualesquiera, sino más específicamente, las expresadas por escrito de manera previa a la vista o a la asamblea. De ello hay ciertas evidencias históricas procesuales pero para M. Kraus resulta una interpretación algo retorcida.

a su definición) y por otro casi omni-abarcantes (en la cantidad de referencias previas que aglutinan) que aquí adquieren el *enthymema* y el *paradeigma.*

En todo caso, podríamos resolver esta especie de conflicto teórico sobre el *ēthos*, su evolución y las distintas perspectivas de Aristóteles, Anaxímenes o Isócrates, de una manera casi trivial, diciendo simplemente que Aristóteles, con su advertencia, está dando instrucciones más bien al orador (aspirante a) y no tanto al auditorio (o a un orador que actúe como oponente). Estaría hablando de aquello que el hablante puede controlar, que sería el *ēthos dia tou logou*, en este caso entendido como el único *ēthos entechnos* posible. Otros elementos de la vida, carácter o acciones del hablante conocidos por su auditorio serían, en este sentido, *atechnoi* y quedarían fuera de su control retórico (lo que coincidiría en cierto modo con la clasificación de la RaA de la *doxa tou legontos*). Por otro lado, la *RaA* trasluciría, por su parte, un exceso de inclinación hacia la controversia (característica del enfrentamiento legal entre partes)[45] más que al "discurso persuasivo", y como tal enfatizaría recursos de tipo más dialéctico (juegos y argumentaciones entre proponente y oponente) que los desplegados por la *Retórica* aristotélica. En ninguno de los dos casos estaríamos hablando de recomendaciones, técnicas o sugerencias para la "evaluación" retórica de argumentos por parte del propio auditorio (que no aparece excesivamente en estos párrafos, precisamente) o de un "analista externo", una figura poco presente en la teoría griega frente a su omni-presencia actual.

Creo, sin embargo, que se puede sacar más jugo de la advertencia aristotélica y de su énfasis en el *ēthos* explícitamente discursivo en términos que interesen a una teoría moderna particularmente centrada en el análisis y la evaluación normativa de argumentos. La pregunta que nos queremos plantear tras esta exploración sería: ¿qué debe hacer un analista/evaluador de argumentos (un analista del discurso) ante los elementos ethóticos del mismo (un analista con una orientación retórica que se interese por tales elementos, se entiende)? ¿Tratar, sobre todo, de centrarse en el acto argumentativo mismo o extraer y explotar información de otras fuentes? ¿Está legitimado el analista para aducir elementos externos (vs. la consideración como falaz del *tu quoque*

[45] Aunque se habla de las referencias legales y judiciales del la *RaA*, hay que destacar que uno de los temas más controvertidos en este texto y de sus relaciones con la *Retórica* de Aristóteles es precisamente su visión y clasificación de los géneros oratorios. Pisani (2011) se centra en analizar, precisamente, el famoso y desconcertante "séptimo género" de Anaxímenes (los otros seis podrían entenderse como variantes positivas y negativas de los tres aristotélicos): el *eksetastikon eidos,* género indagatorio o investigativo. Precisamente los rasgos que destaca Pisani en este género (a partir de oraciones contemporáneas) tienen mucho interés para el tema que nos ocupa, mencionando entre otras cosas que el género incluiría: a) búsqueda de las contradicciones de la persona investigada entre intenciones, acciones, discurso, elecciones, alianzas, contraste con anteriores actitudes etc.. (supuestamente para construir *tekmeria*); b) búsqueda de oposiciones entre las acciones, actitudes del investigado y la moral común o aceptada o c) discrepancias respecto de ciertos valores o fines como la justicia, legitimidad, el interés común o la plausibilidad (razonable) (supuestamente para construir *enthymemata*).

u otras alegaciones *ad hominem*)? ¿Está legitimado el argumentador para obviar tales elementos y pedir una recepción "cándida" de su discurso?

No se trata, evidentemente, de opciones excluyentes y la conveniencia de seguir uno u otro principio dependerá de los contextos y prácticas argumentativos involucrados, pero conviene reflexionar sobre el sentido de lo recomendado por Aristóteles (una opción que podría resultar a la vez cara a los más lógicos por su restricción al texto y rechazada por sus aspiraciones más epistémicas). Se trata, en principio, de una opción menos fácil de seguir, menos "natural" para un receptor real de un discurso (como lo demuestran ciertos casos reales: e.g. el *escándalo* de F. Rico, 11/01-06/03 de 2011) y que obliga a una actitud intelectualmente problemática, aparentemente poco justificable en términos epistémicos o de relevancia de la "información verdadera" sobre "estados de cosas". Pero se trata de un "ejercicio" (si se quiere) de una presunción táctica que, sin embargo, permite la apertura de ciertas posibilidades que son relevantes desde el punto de vista argumentativo y comunicativo. Mantener una inicial presunción sobre la pertinencia exclusivamente interna de tipo *ethótico* del discurso no solo permite al auditorio seguir el imperativo ético de dar una oportunidad renovada (independiente de su "vida anterior") al hablante en su cometido de dar cuenta y razón de su postura en el momento de la comunicación, sino que además abre la posibilidad al orador de representar intereses y voluntades ajenos o solo parcialmente propios o poner de relieve cuestiones sobre la pertinencia y los derechos de participación de las diversas voces involucradas en una controversia. De nuevo, el "escándalo Rico", puede servirnos de ejemplo para algunos de estos rendimientos argumentativos.

REFERENCIAS

Amossy, R., ed. 1999, Images de soi dans le discours. La construction de l'ethos, Ginebra, Delachaux et Niestlé.

Amossy, R. 2006 [2000], L'argumentation dans le discours, París, Armand Colin.

Amossy, R. 2009, 'La double nature de l'image d'auteur', Argumentation et Analyse du Discours 3, on-line, consultado el 06/03/2012, http://aad.revues.org/662.

Anaximenes de Lámpsaco 1989, Retórica a Alejandro, trad. J. Sánchez Sanz, Salamanca, Universidad de Salamanca.

Aristóteles, 1985, Retórica, ed. y trad. Antonio Tovar, Madrid, Centro de Estudios Constitucionales.

Bokobza Kahan, M. 2009, 'Introduction. Ethos discursif et image d'auteur', Argumentation et Analyse du Discours 3, on-line, consultado el 07/03/2012, http://aad.revues.org/658.

Booth, W.C. 1961, The Rhetoric of Fiction, Chicago: Chicago University Press.

Chiron, P. 2011, 'Relative Dating of the *Rhetoric to Alexander* and Aristotle's *Rhetoric*: A Methodology and Hypothesis', Rhetorica XXIX/3, 236-262.

Kraus, M. 2011, 'How to Classify Means of Perusasion: The *Rhetoric to Alexander* and Aristotle on *Pisteis*', Rhetorica XXIX/3, 263-279.

Mainguenau, D. 2009, 'Auteur et image d'auteur en analyse du discours', Argumentation et Analyse du Discours 3, on-line, consultado el 06/03/2012, http://aad.revues.org/660.

Meizoz, J. 2009, 'Ce que l'on fait dire au silence : posture, ethos, image d'auteur', Argumentation et Analyse du Discours 3, on-line, consultado el 07/03/2012, http://aad.revues.org/667.

Noël. M.-P. 2011, 'Isocrates and the *Rhetoric to Alexander*: Meaning and Uses of *Tekmerion*', Rhetorica XXIX/3, 319-335.

Pisani, G. 2011, 'The *ekstastikon eidos* of the *Rh.Al.* and Parallels in Aeschines' *Against Timarchus* and Demosthenes' *On the False Embassy*', Rhetorica XXIX/3, 336-356.

LA ARGUMENTACIÓN EN LA ESFERA PÚBLICA DEL DISCURSO*

Coordinador LUIS VEGA REÑON
UNED

En los estudios sobre argumentación hay un área transdisciplinar que viene concitando mayor atención y desarrollos específicos desde las dos últimas décadas del pasado siglo: es el área abierta por la consideración de la llamada "esfera pública del discurso", donde confluyen motivos e intereses varios y diversos (e.g. analíticos y socio-institucionales, cognitivos y retóricos, éticos y políticos). Como prolongación y concreción del simposio anterior, en dicha área, este simposio debatirá cuestiones relacionadas con la deliberación desde distintos puntos de vista, con nuestros mecanismos argumentativos y con los nuevos recursos y medios de comunicación discursiva.

¿ES POSIBLE MEDIR LA CALIDAD DE LA ARGUMENTACIÓN EN LA ERA DIGITAL?

ENRIQUE ALONSO Y MIGUEL VIVANCO
Universidad Autónoma de Madrid
Enr.alonso@gmail.com

ABSTRACT: Social Networks have created new phenomena in the field of argumentation that is necessary to study. Of all the enviroments, this work focuses on social networks and digital media. Both are strongly related through the user's contributions. This involves designing automated monitoring tools for these argumentative environments and analyzing the way in which the obtained data should be interpreted.
KEYWORDS: Monitoring the Web; Argumentation in the Web; Social Networks

1. Un clima favorable

La desconfianza ante la capacidad de las tecnologías de la información para colaborar de forma activa en los procesos de transformación social ha sido un tópico defendido con fe desde las filas de la tecnofobia más militante. Toda nueva tecnología debía suponer antes o después el origen de nuevas penalidades y un nuevo paso atrás desde alguna supuesta *Arcadia* de la que cada vez nos alejábamos más.

* Desarrollado en el marco del proyecto financiado por la Secretaría de Estado de Investigación, desarrollo e Innovación, FFI2011-23125.

El uso reciente de estas tecnologías en movimientos populares de largo alcance -*La Primavera Árabe, Los Indignados, Occupy Wall Street, etc*- parece apuntar en otra dirección haciendo que por fin podamos dedicar un tiempo a su estudio sin parecer por ello defensores de alguna distopía futurista. Las nuevas tecnologías han creado usos y formas de acceso a la información que han calado en grandes sectores sociales, quizá más en aquellos que por primera se sienten capaces de participar activamente en procesos de creación de noticias sin depender de quienes hasta ahora habían disfrutado de su monopolio. En ese sentido parece posible hablar de la aparición de *sistemas horizontales de difusión de la información* como algo opuesto al modelo vigente hasta hace bien poco. Los *modelos verticales*, agrupados en torno a grandes gestores de información que actúan como referencias exclusivas, es el representado por las grandes empresas que dominan los sectores de la radiodifusión tradicional, la Televisión, la prensa escrita o las editoriales convencionales. Es desde sus filas y todo el entramado que les rodea, desde donde más críticas se pueden oír a las nuevas formas de propagación de la información y no es de extrañar, ya que su adaptación a las nuevas condiciones no está resultando fácil en ningún caso.

Pero el nuevo modelo tampoco representa en sí mismo la solución a todos los problemas de acceso de los ciudadanos a la información y sobre todo, dista aún mucho de ser un marco que realmente favorezca el debate público y la buena lógica civil. Lo único que ha cambiado es el clima previo de escepticismo roto por el uso de las nuevas tecnologías en movimientos que, al menos en su programa, defienden opciones de progreso muy alejadas de las distopías tecnológicas típicas.

En este trabajo nos planteamos la posibilidad de analizar con técnicas basadas en el funcionamiento real de la Red el papel del debate en lo que llamaremos *los nuevos ambientes de la argumentación*. Somos conscientes de la existencia de dos grandes obstáculos:

1. La ausencia de un consenso suficiente acerca de qué debe constar como una argumentación y
2. La dificultad técnica para identificar estructuras que nos permitan pensar en la presencia actos de argumentación en distintos medios digitales.

La tensión entre ambos extremos se acrecienta si tenemos en cuenta la dificultad existente a veces para diferenciar entre lo que es la mera expresión de una opinión, o una adhesión poco o nada reflexionada, con un genuino debate. La celeridad con que ciertos procesos de manifestación activa de la propia posición se pueden llegar a producir en determinados medios digitales crea una suerte de espacio intermedio entre la mera calificación de contenidos y su réplica argumentada que no siempre es fácil de identificar, pero esto es a lo que nos enfrentamos. Y debemos intentar al menos

hacerlo de forma práctica, de tal manera que nuestras hipótesis puedan ser evaluadas sobre el inmenso campo de estudio que hoy por hoy representa la Red[1].

2. Nuevos ambientes para la argumentación

Intentar ser exhaustivo en este punto no parece conveniente en absoluto. La vigencia de las herramientas que en cada momento resultan populares no está garantizada más allá de unos pocos años, por lo que hacer teorías de gran recorrido no resulta indicado, simplemente pueden volverse obsoletas antes de darse a conocer. No obstante, sí parece adecuado hacer un repaso por lo que hasta ahora han sido los principales ambientes para la comunicación en la era digital. La lista debería contener al menos los siguientes *items*:

- Correo electrónico
- Chat
- *Posteo* de contenidos en Redes sociales
- El comentario de contenidos en medios digitales

En esta ocasión nos centraremos en los dos últimos componentes de la lista. La razón no es la ausencia de interesantes fenómenos argumentativos en ellos, sino la imposibilidad de someter estos ambientes a una monitorización apropiada.

Quizá sea el momento de aclarar que cuando hablamos de *monitorización* hacemos referencia al uso de herramientas digitales que permitan seguir y analizar de forma automática los datos requeridos. Este término ha adquirido cierta popularidad a partir de la aparición de recursos que permiten imitar la tecnología empleada por los motores de búsqueda de herramientas tipo *Google*. Cuando analicemos más adelante la posibilidad de monitorizar un determinado ambiente para la argumentación, lo que se estará valorando es la existencia de aplicaciones capaces de llevar a cabo ese cometido.

2.1 *Posteo* de contenidos en Redes sociales

La denominación *redes sociales* hace referencia en realidad a un número muy reducido de entornos. De hecho, podríamos centrarnos en tres de ellos sin especial problema: *Facebook*, *Twitter* y *Youtube*. Se trata de ambientes muy diferentes en los que está surgiendo un cierto tipo periodismo, si así se le puede llamar, basado en el comentario de contenidos generalmente copiados o reproducidos desde otros medios. En esta ocasión nos centraremos en los dos primeros.

[1] . Mientras la mayor parte de la Red siga funcionando con las tecnologías vigentes basadas en HTML y otros lenguaje de código abierto, esta será un entorno en el que podremos realizar análisis e investigaciones razonablemente transparentes. Si eso cambia, y hay indicios de que lo está haciendo, la cosa puede ser muy distinta

Un tipo de comportamiento completamente habitual en ambos entornos consiste en la elaboración de contribuciones –*post*– que constan de un enlace a un recurso externo con un pequeño anticipo –*titular*– al que se añade un comentario más o menos extenso. Originalmente, este tipo de contenidos suponían un cierto esfuerzo, ya que ni *Facebook* ni *Twitter* habían previsto un uso tan recurrente de este tipo de acciones. Dicho de otra forma, el usuario fue en esto, como en tantas otras cosas, muy por delante de los diseñadores. Más tarde vendrían los célebres *botones* de *Facebook, Twitter*, etc que facilitan, desde el propio origen, una forma fácil de enlazar y comentar el contenido deseado.

Estos *posts* o contribuciones se convierten así en un género novedoso en el que el usuario inicia un debate con su entorno a partir de un comentario a un contenido ajeno. Ese entorno es limitado y cerrado en *Facebook* y mucho más abierto, público y social en *Twitter*, pero el comportamiento básicamente es el mismo.

¿Estamos ante un caso genuino de argumentación, es un género comunicativo nuevo? La respuesta a la primera pregunta parece claramente afirmativa. El usuario que actúa de esa forma adopta la actitud de quien se pronuncia a favor o en contra de una tesis –representada de forma más o menos explícita en el contenido original reproducido– que apoya con razones mediante su comentario. El resto de los participantes suelen aportar sus razones provocando debates de longitud variable. Otra forma de contribución tiene lugar mediante el *equívoco* botón de *me gusta*. La pregunta acerca de si estamos ante un género nuevo es más compleja de resolver. Este ambiente adopta parte de la estructura característica de la propagación del *rumor* y esta es la razón de que se emplee a menudo en procesos de comunicación viral. A parte de eso, es cierto que es el usuario el que decide actuar sobre un contenido en un entorno que considera de su propiedad, lo que habitualmente se denomina un *perfil*. Esta dimensión por la que el usuario se convierte de manera explícita en un actor de los flujos de información sí parece novedosa.

A parte del indudable impacto que estas redes han tenido y tienen en procesos sociales, tanto cotidianos como extraordinarios, lo interesante del caso es que sí son monitorizables, al menos en una medida que permite imaginar el diseño de programas interesantes de análisis y seguimiento del debate.

2.2 El comentario de contenidos en medios digitales

En esta ocasión se trata de fomentar el debate de manera explícita. La diferencia es que la noticia y el comentario no se propagan del modo que lo hacen en el caso anterior, sino que quedan contenidos dentro del medio que lanza la noticia original, normalmente un medio de comunicación

reconocido y con una presencia digital estable. Es decir, no existe el efecto viral por el que tanto se han preocupado los teóricos de la información en los últimos años.

A diferencia de las redes sociales, los comentarios en prensa digital están abiertos a comunidades con poco o ningún control, lo que suele hacer que los debates degeneren a menudo y deban ser moderados.

Tanto aquí, como en el caso de las redes sociales, interesa analizar la extensión de los debates, los participantes que intervienen como promedio, las formas en que deben ser moderados y las veces que se exporta un contenido hacia una red social. Pero hay muchos otros elementos que pueden ser analizados gracias al hecho de que este ambiente es ampliamente monitorizable y está, además, muy fuertemente conectado con el anterior.

3. ¿Es posible en general monitorizar el debate?

Esta pregunta tiene dos posibles interpretaciones. En primer lugar, puede y de hecho hace referencia a un problema técnico: ¿permite la estructura de la Red construir aplicaciones útiles en este sentido? Pero también tiene una inevitable vertiente conceptual: ¿existen herramientas teóricas adecuadas para el análisis de este tipo de nuevos ambientes para la argumentación?

En relación al primer problema hay que aclarar que no se trata de saber si existen o no medios para medir la actividad de los usuarios en las Redes sociales y los medios de comunicación digitales, porque es obvio que existen. Los propietarios de esos medios poseen y de hecho emplean con intensidad recursos automáticos que constantemente monitorizan nuestra actividad en sus sitios. Más bien se trata de saber si nosotros también podemos acceder al menos a una parte de esa información. En los últimos años se ha puesto a disposición de desarrolladores y expertos una serie de herramientas dentro del movimiento del *software libre* que permiten una captura significativa de los datos de ciertos sitios sin aplicar técnicas intrusivas. Su uso se limita a medios basados en *html* por lo que quedarían fuera de su alcance nuevos medios de comunicación y debate centrados en dispositivos ligeros tipo *smartphone*. Tanto para ciertos datos internos de las redes sociales como para comunicaciones orientadas expresamente a *smartphones*, mucho me temo que deberemos confiar en lo que los gestores y propietarios de estos entornos nos quieran decir, pero hay un espacio considerable en el que esto aún no es necesario. Es ahí donde se enmarca este trabajo y el diseño de herramientas apropiadas.

El segundo problema es con mucho, el más complejo de todos. Dejémoslo por el momento en algo mucho más modesto, una simple lista de aquellas cosas que nos gustaría saber, con las que obtendríamos información valiosa para entender el comportamiento de la argumentación en los nuevos ambientes.

- Análisis y seguimiento de tendencias de los *trending topics* de *Twitter*.
- Análisis de la longitud, número de intervinientes y niveles de moderación en los debates en prensa digital.
- Análisis de ese mismo hecho en las redes sociales con especial atención al uso de *botones de valoración* –me gusta, etc–.
- Uso de *hashtags* y otros términos clave como forma de fijar y orientar un debate.
- Propagación de los debates y noticias aparecidos en prensa digital hacia los perfiles de usuario en redes sociales.

Por el momento esto parece más que suficiente. Creo que debemos tener en cuenta, y valga como resumen de lo que hemos querido decir aquí, que el estudio de los ambientes digitales en que ahora se desarrolla la comunicación y el debate no puede limitarse al uso de técnicas tradicionales. Hay aspectos que el empleo de técnicas de monitorización nunca –aunque eso tal vez es mucho tiempo– van a poder tratar adecuadamente. Pensamos en todos aquellos que tienen que ver con el *contenido* de la argumentación, pero en todo lo demás sí es posible servirse con ventaja de las técnicas digitales disponibles. No hacerlo, no intentarlo al menos, no tendría justificación en este momento.

REFERENCIAS

Cremades, J.: Micropoder. La fuerza del ciudadano en la era digital. Espasa-Calpe, Madrid (2007)

Fernández de Lis, P.: "El ciberespacio se come al libro", El País (2007). <http://www.elpais.com/articulo/sociedad/ciberespacio/come/libro/elpepusoc/20070307elpepisoc_2/Tes >

González Barahona, J.M.: El futuro de la información: ¿vamos a donde queremos? <http://www.sinetgy.org/jgb/articulos/futuro-informacion/> [Consultado 12 noviembre 2011]

Librería cliente URL.: <http://php.net/manual/es/book.curl.php> [Consulta:1 de mayo de 2012]

Stallman, R.: ¿Quién vigilará a los vigilantes? < http://biblioweb.sindominio.net/telematica/watchmen_es.html> [Consulta: 24 abril 2011]

Twitter Development.: <https://dev.twitter.com/> [Consulta:1 de mayo de 2012]

MECANISMOS RETÓRICOS ARGUMENTATIVOS Y NO-ARGUMENTATIVOS

LILIAN BERMEJO LUQUE
Universidad de Granada.
Lilian.bermejoluque@gmail.com

KEYWORDS: Rhetorical dimension, argumentative and non-argumentative rhetorical devices, public discourse

La concepción tradicional de la dimensión retórica de la comunicación como proceso de inducción simbólica plantea la cuestión de hasta qué punto sea posible identificar determinados efectos perlocutivos como efectos genuinamente retóricos. Mediante el análisis de algunos ejemplos del discurso político, trataremos de dar respuesta a esta dificultad y propondremos una distinción entre mecanismos retóricos argumentativos y no-argumentativos.

DELIBERACIÓN Y COGENCIA ARGUMENTATIVA

JOSÉ M. SAGÜILLO
Departamento de Lógica y Filosofía Moral
Universidad de Santiago de Compostela.
josemiguel.saguillo@usc.es

ABSTRACT: This paper investigates the ways in which deliberation takes place in the practice of argumentation, namely in the realm of argumentative practice. The discussion is focused on the whereabouts taken place in the mind of the thinker when constructing cogent pieces of reasoning leading from premises to conclusion. It is shown that the epistemic enterprise of arguing requires making choices and presupposes rational deliberation facing available evidence.
KEYWORDS: argumentation, practice, deliberation, cogency, rational choices, epistemics

1. Introducción

La palabra 'deliberación' no sólo es una expresión ambigua con diferentes significados en distintos contextos, sino que además, dado uno de sus sentidos, es posible identificar distintas clases de deliberación y modos de deliberar que caerían bajo dicho sentido. 'Deliberar' proviene del latín 'deliberare' y significa en general ponderar, o medir atentamente, los pros y los contras de algo antes de tomar una decisión. Contiene el étimo 'libra' que significa balanza; de ahí la metáfora recurrente de que deliberar monta tanto como "poner en la balanza", "pesar" o "sopesar" distintas opciones a efectos de decidir a favor de la mejor.

En una de sus acepciones, la deliberación se entiende vinculada a la toma de decisiones sobre asuntos *prácticos*. En este trabajo el objeto de la deliberación también incumbe a la toma de decisiones sobre asuntos no-prácticos. En particular, el presente es un estudio de la deliberación que se exhibe en la práctica argumentativa (no necesariamente en la argumentación práctica). Distinguimos entre la deliberación sobre asuntos prácticos y la práctica de la deliberación, donde la práctica de la deliberación puede dirigirse a cuestiones prácticas y a cuestiones teóricas. Así, en la presente aproximación a la deliberación se toma como fuente de estudio la práctica argumentativa y no la argumentación práctica, entendiendo que la deliberación tiene un papel en el propio proceso argumentativo que es previo al juicio de verdad y cogencia. Sin duda un análisis tal de la

deliberación que opera en la práctica de la argumentación requerirá incorporar no sólo elementos monológico-individuales sino también elementos dialógico-colectivos.

En este enfoque de la deliberación centrado en la práctica argumentativa es importante notar que la propia deliberación *es* un medio y *es* una condición necesaria para decidir en el propio proceso de argumentar. Cuando la argumentación es un medio para la deliberación, es decir, cuando la argumentación proporciona elementos de juicio para sopesar, propongo emplear la expresión *argumentación deliberativa*, mientras que para referirnos a la deliberación a efectos de configurar o construir una argumentación, propongo emplear la expresión elíptica *deliberación argumentativa* como abreviatura de deliberación, sea monológica, sea dialógica, en la práctica argumentativa. Nótese pues que el estudio de la deliberación argumentativa, el asunto de este ensayo, señala al papel de la deliberación para la toma de decisiones en el proceso de elaboración de una argumentación, y que es previo al juicio sobre la cogencia de nuestras argumentaciones.

El sentido que se propone del estudio de la deliberación en la práctica argumentativa es coherente con la propia visión de Aristóteles sobre este tema, tal como se expresa en su Ética, cuando dice que la deliberación atañe "a las cosas que dependen de nosotros y que son realizables" (1112a 30-31). Sin duda, la deliberación específica con un rol clave en el proceso o "acción" de argumentar es algo que nos atañe de modo esencial. Como correctamente indica Luis Vega (2010), la discusión sobre la deliberación del estagirita tiene un dibujo de lo conveniente en su Retórica y tiene un dibujo prudencial en su Ética. Creo sin embargo importante recordar que la deliberación también está presente en la práctica de lo que recogen sus estudios teóricos de los *Primeros Analíticos*, aunque Aristóteles no fuera explícito al respecto, puesto que la cuestión por la deliberación emerge también de modo natural con un dibujo epistémico o cognoscitivo en el ámbito de la propia génesis de la lógica y de la cogencia argumentativa.

2. La deliberación argumentativa

Deliberamos con respecto a cada una de las partes de una argumentación P,C,R, formada por un conjunto de premisas P, una cadena de razonamientos intermedios R y una conclusión C, antes de tener una aproximación a las "piezas concretas" que de facto formarán parte de su realización. La argumentación no existe como tal hasta la completa determinación de cada una de sus partes, premisas, cadena de razonamiento intermedio y conclusión, no necesariamente en este orden.

En la práctica, cierta situación y cierto contexto de problematicidad, señalará o aislará de modo deliberativo una cierta proposición crítica o hipótesis, es decir, una proposición que no se sabe verdadera ni se sabe falsa. La deliberación con respecto a dicha hipótesis y quizás a un curso de acción asociado, involucra sopesar si la hipótesis es o no es acertada. La determinación

"deliberada" de que en definitiva se trata de una buena idea o de que en definitiva se trata de una mala idea, se obtiene o se justifica –en los casos que nos ocupan—mediante argumentación.

Por un lado, el trabajo deliberativo para determinar que la hipótesis es acertada requiere a su vez de la deliberación tentativa-cautelar con respecto a qué conjunto de premisas contendría la información suficiente para que exista la relación de "seguirse de", entre la hipótesis y el conjunto de premisas. También requiere por descontado, de un razonamiento justificador o de una inferencia "sensata" ajustada a la situación, al contenido de la hipótesis y al de las premisas.

Por otro lado, la determinación "deliberada" de que una hipótesis no es acertada también se realiza y se justifica mediante argumentación. Esencialmente el trabajo deliberativo previo a la obtención de dicha argumentación requiere a su vez deliberar sobre qué conjunto de premisas consideradas como buenas ideas, unido a la hipótesis considerada como no acertada, implicaría, en algún sentido de 'implicar', a una conclusión que ya sea reconocida o sabida previamente como desacertada, como una conclusión falsa o indeseable.

En muchos casos se podría objetar que no deliberamos cuando procesamos información deductiva para determinar relaciones de consecuencia mediante pasos de inferencia elementales, quizás ya sabidos a priori como cogentes. Sin embargo, en muchos casos el procesamiento de información en la práctica deductiva involucra pasos de inferencia que distan de ser obvios, sin mencionar la relatividad al participante de lo que es obvio y de lo que no lo es, de lo que se juzga como razonamiento cogente y de lo que no. El hecho de que Aristóteles nos indicara que no deliberamos con respecto a lo necesario sino con respecto a lo contingente no debe enmascarar que en realidad sí deliberamos con respecto a lo que *no juzgamos* necesario aunque pudiera serlo, y que también deliberamos con respecto a lo que juzgamos contingente aunque de facto pudiera ser necesario, y éstos son precisamente los casos de estudio interesantes de la práctica inferencial.

En el contexto de la discusión acerca de la distinción entre inferencia inmediata y mediata es importante la siguiente pregunta: ¿Deliberamos y decidimos para saber que **P** implica **c**, si de facto **P** implica **c**, o más bien se nos impone que **P** implica **c**, si de facto **P** implica **c**? Resulta iluminador plantear la pregunta análoga para el caso complementario opuesto: ¿Deliberamos y decidimos para saber que **P** no implica **c**, si de facto **P** no implica **c**, o más bien se nos impone que **P** no implica **c**, si de facto **P** no implica **c**? Conviene notar que 'imponer' es un término normativo, más propio del ámbito de la jurisprudencia, y que 'proponer' es un término deliberativo propio del contexto del descubrimiento deductivo que debería naturalmente propiciar (como algo derivativo) el consenso o el disenso. De modo ponderado y cualificado, se mantiene que en un contexto genuino de descubrimiento deductivo deliberamos para "captar" y "enjuiciar" si es el caso o no es el caso que **P** implica **c**, aun en los casos más inmediatos. Se sugiere que virtualmente que no hay ningún juicio de

validez o de invalidez calificado como "obvio", "visible" o "evidente", que no haya tenido un sujeto real para quien el esfuerzo de "llegar a ver" la relación de implicación o de independencia no fuera difícil, donde no se sopesaran evidencias, y donde no se deliberara individual o colectivamente para saber cuál es el caso. Nótese que esto es compatible con el hecho de que una vez que "vemos" que P implica c, también nos percatamos de que no podría ser de otra manera. Lo importante aquí es no olvidar que deliberamos y hacemos cierto esfuerzo intelectual responsable para llegar a juzgar que P implica c, o que P no implica c.

La comprensión de la evidencia y de la información disponible; la ponderación sobre consecuencias deseadas y no deseadas; la subsecuente re-evaluación sobre la posible insuficiencia de la información disponible, sea fáctica o sea inferencial, y su recolección ulterior que dan lugar a cualificar una subclase de consecuencias deseadas que neutralicen las indeseadas, son todas intuiciones que sugieren, entre otras cosas, que la práctica de la deliberación argumentativa parece requerir en muchos casos principios no-monótonos de revisabilidad, al menos para algunos de sus estadios iniciales. En nuestra red de creencias frecuentemente empleamos más de un relación de "seguirse de", la que sea más propicia dada lo que sabemos o lo que aceptamos, tanto a la hora de determinar si tanto la información disponible como su procesamiento es suficiente, y por supuesto, si dicha información es en efecto "fidedigna" o "buena".

Comprobamos también que en un estadio maduro del proceso argumentativo, cuando la relación de "seguirse de" sea lógica o analítica, los dos movimientos descritos previamente, es decir la búsqueda de implicantes juzgados como verdaderos o la búsqueda de implicaciones juzgadas como falsas, no son más que descripciones epistémico-fenomenológicas de realizaciones exitosas del método deductivo, las cuales nos proporcionan pruebas, y de las realizaciones exitosas del método hipotético-deductivo, las cuales nos proporcionan sólo deducciones.

La práctica de la deliberación argumentativa en muchos casos se resuelve pues en una especie de relación de *precedencia* con la búsqueda "ascendente" de implicantes o de *sucesión* con la búsqueda "descendente" de implicaciones. En ambos casos, en el inicio del proceso que se trate, la argumentación es mera intención, puede que ni siquiera esquemática, debido a la disposición material de sólo alguna de sus partes, la cual adopta a las luces del agente, el rol de guía en la solución "deliberada" de los "huecos", factores o incógnitas que permanecen abiertas o por resolver. Existe por tanto una relación de *feedback* que se establece entre la deliberación sobre cada uno de los factores y la deliberación factorial final de la argumentación obtenida. Ya Descartes indicaba la posibilidad real de que un sujeto pensante constatara la cogencia de cada uno de los pasos de razonamiento en una argumentación dada pero que careciera de la constatación de la cogencia global de todo el recorrido inferencial.

3. Estructura, límites y errores de la deliberación argumentativa

En una primera aproximación, no deberíamos perder de vista que deliberamos para tomar decisiones acertadas. En el análisis del proceso de toma de decisiones la relación de preferencia y la relación de indiferencia de un sujeto con respecto a un universo de opciones tienen un rol descriptivo-extensional central. Desde un punto de vista lógico formal, la relación de preferencia [estricta] de un sujeto x sobre un universo U de opciones es irreflexiva, asimétrica y transitiva, y la relación de indiferencia es reflexiva, simétrica y transitiva, es decir, de equivalencia. Podríamos decir que si un sujeto no tiene internalizadas *a priori* las propiedades abstractas de sus relaciones de preferencia y de indiferencia, dicho sujeto no sería racional, lo que lo haría incapaz de deliberar. Diríamos que un sujeto tal, ni siquiera "entra en el juego de las balanzas" o que "sus balanzas no están adecuadamente calibradas para comparar pesos *objetivos*".

Combinando las dos relaciones previas de preferencia e indiferencia obtenemos la condición de la tricotomía: dado un universo de opciones, un sujeto ideal x determina que para dos opciones *a* y *b* cualesquiera, exactamente se da uno de los siguientes casos: *a* es preferible a *b*, o *b* es preferible a *a*, o *a* y *b* son indiferentes. Desde un punto de vista material, y atendiendo a las capacidades intelectuales de sujetos-personas (no ideales), hay muchos universos en los cuales la tricotomía no es epistémicamente accesible. En efecto, cuando carecemos de conocimiento y experiencia sobre un asunto, o cuando el universo de opciones es demasiado amplio, podríamos no ser capaces de "tricotomizar" dicho universo. El hecho es que de saber más ampliaríamos nuestros criterios deliberativos y decidiríamos mejor.

Las modernas teorías de la decisión parecen sustentarse en ideas que tienen una venerable tradición en la epistemología: los dos paradigmas que podrían ser eventualmente convergentes se basan en la idea del equilibrio del sistema razonador y en la idea de la maximalización de la utilidad esperada. Es decir, estas teorías presuponen que la toma de decisiones se sustenta en el principio de "exprimir" nuestra red de creencias y sus interconexiones inferenciales a efectos de purgarla de lo falso y de expandirla con lo verdadero, en ambos casos, con vistas a potenciar predicciones y líneas de acción exitosas.

Lo que estas teorías de la decisión no incorporan en su discurso—ciertamente de modo análogo al que las teorías lógicas no incorporan muchos errores de la práctica argumentativa—es que a menudo deliberamos mal debido a nuestros propios sesgos, debido a lo que podríamos llamar la carencia de un "buen" sentido de la proporción. No se trataría en este caso tanto de un error de calibrado formal, ni tan siquiera de una limitación material en sentido estricto, sino más bien de no saber emplear bien nuestra balanza con la debida energía y con la debida prudencia. La cuestión en definitiva es no fantasear o no enmascarar la evidencia y afinar nuestras capacidades intelectuales

en sus aplicaciones, es decir, "mirar para ver"; "escuchar para oír"; "meditar para comprender e inferir", potenciando el uso del entendimiento formal y su realización material óptima mediante el ejercicio deliberativo.

En un sentido general, en muchos casos, estos errores deliberativos pueden considerarse falacias de proceso-producto entre cada una de las partes de una argumentación con respecto al resultado final. Nuestra fragilidad razonadora es notoria. Conviene recordar en este punto que el binomio ignorancia-error, error-ignorancia es a menudo responsable del "hallazgo" de paradojas en los intentos de la expansión argumentada del conocimiento; del intento de descubrimiento inferencial sustentado en supuestas relaciones de consecuencia. En estas circunstancias el valor de la deliberación argumentativa se muestra en acción con toda su estampa característica. La obtención de una creencia inaceptable a través de un razonamiento creído como cogente a partir de creencias aceptadas nos mueve necesariamente a suspender nuestras creencias y a re-examinarlas. En este escenario, deliberamos para poder comenzar a re-examinar nuestras creencias, estableciendo una relación de orden tentativo reconsiderando premisas, o reconsiderando el razonamiento o reconsiderando la conclusión, en el orden que el estado epistémico y la deliberación previa veces del sujeto argumentador le sugiera. Un cambio de creencia supone que la paradoja "se solventa" y queda re-clasificada o como *candidata* a deducción en el primer caso, o como *candidata* a falacia en el segundo, o como *candidata* a prueba en el tercero. El procedimiento de resolución final de la paradoja requiere de ulterior evidencia y deliberación hasta llegar a saber cuál o cuáles de nuestras creencias previas erran incorrectas.

Observaciones finales

Hay estados epistémicos completos con respecto a cierto universo o dominio de investigación restricto pero todo estado epistémico es necesariamente parcial e incompleto con respecto a un universo suficientemente irrestricto. Esto forma parte esencial de la empresa dinámica de llegar a saber y tenemos que vivir con ello deliberando a cada paso. Es saludable recordar en este punto que el paradigma lógico dominante del siglo XX "expulsó" al sujeto deliberativo del escenario metodológico. Deliberamos para poder decidir sobre la mejor base epistémica posible y teniendo presente el peligro latente de nuestros propios sesgos observacionales e inferenciales. Nunca ha sido de otro modo. Si aislamos lo que de dogma y mito a veces se genera en una disciplina suficientemente madura veremos que la práctica de la medicina, como la práctica del derecho, o como la práctica matemática están todas ellas sometidas a nuestras capacidades intelectuales en pugna por "reducir" una realidad externa, problemática y siempre exigente. Nuestras mejores teorías y tratados de medicina, de derecho, y como no, de argumentación, intentan

reflejar lo que dichas prácticas han "sancionado" como buenas prácticas. La ubicuidad de la acción intelectual de deliberar es también otra baza interesante para aquilatar cierta brecha que se ha producido entre el estudio de la argumentación matemática y el estudio de la argumentación jurisprudencial, que se supone ponderada y atenta a las circunstancias más o menos contingentes que rodean el caso bajo consideración. Hemos intentado mostrar que al menos en lo que a la acción de deliberar se refiere, ambos tipos de prácticas argumentativas requieren ponderación y aquilatamiento de evidencia, y que la apelación a las contingencias que hacen al caso, parece admitir a lo sumo, una diferencia de grado.

Agradecimientos: Este trabajo se realiza bajo la cobertura de los proyectos FFI2011-23125, y FFI2009-08828 del Ministerio de Ciencia e Innovación. Debo mi agradecimiento a Luis Villegas, Concha Martínez, Luis Vega, Ramón Jansana, Javier De Donato, Julio Armero, Eduardo Bustos, Paula Olmos, Jesús Zamora y Manuel Pérez por comentarios e intercambios sobre una versión previa.

REFERENCIAS

Aristóteles: *La Scienza della Prassi. Da Ética Nicomachea e Política*. A cura di G. Stelli. Armando Editore: Milano 2009

Aristóteles: *Retórica*. Introduzione di Franco Montanari. Testo crítico, traduzione e note a cura de Marco Dorati. Oscar Mondadori: Milano 1996.

Borobej, M. 2006: *A Theory of Argument*. Cambridge: Cambridge University Press.

Cavender, Nancy y Kahane, Howard: *Logic and Contemporary Rhetoric*. Wadsworth: Belmont, CA, USA 2006/10.

Corcoran, J. (1989): 'Argumentations and logic'. *Argumentation* 3, 17-43

González Serrano, Urbano: *Manual de Psicología, Lógica y Ética*. Volumen II Lógica. Gregorio Hernando: Madrid 1945.

Govier, T. 2010: *A Practical Study of Argument*. Belmont, CA: Wadsworth.

Sagüillo, J. 2006: Argumentative practice: cogency and normativity. *Actas del VII Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España* Granada: Universidad de Granada.

Skyrms, Brian: *The Dynamics of Rational Deliberation*. Harvard University Press: Cambridge Massachusetts & London, England 1990

Vega, Luis: ¿Qué debe saber de lógica un (buen) ciudadano? XIII EIDL, Morelia 2010.

Zamora, Jesús: Three types of deliberation. Presentación no publicada.

ELOGIO DE LA RETÓRICA*

LUIS VEGA REÑÓN
Dpto. de Lógica, Historia y Filosofía de la Ciencia, UNED.
lvega@fsof.uned.es

ABSTRACT: I intend to highlight the relevance of the deliberative rhetoric to the public sphere of discourse from a classical reference such as represented by Aristotle's *Rhetoric*.
KEYWORDS: Aristotle, Rhetoric, rhetorical argumentation, public discourse.

1.

De entrada, me he tomado la licencia de un título retórico: puede ser provocador, si no resulta impertinente, en un Congreso de una Sociedad de Lógica y Filosofía de la Ciencia. Desde luego, busca oponerse a una suposición por lo regular tácita pero extendida entre los lógicos y filósofos de la ciencia que oyen hablar de retórica: el supuesto de la partición del campo del discurso en el par de opuestos lógica vs, retórica, de acuerdo con el axioma:

1. La razón es a la lógica lo que la seducción a la retórica.

De donde se deriva por "alternancia" [2] la conclusión:

2. La razón es a la seducción como la lógica es a la retórica,

que sentencia su exclusión mutua.

Esta partición descansa en una idea sesgada de la retórica, que aún se mantiene como si el tiempo, desde mediados del pasado siglo hasta hoy, hubiera transcurrido en vano [3]. La verdad es que uno de los legados reconocidos de la 2ª mitad del siglo XX ha sido la constitución de la retórica como una nueva o renovada disciplina dentro del campo de la argumentación, más allá de la tradición escolar que la reducía a la oratoria o a un género ornamental literario o estilístico y, por cierto, al margen de su inclusión entre las malas artes de la comunicación de masas. En nuestros

* En el marco del Proyecto FFI2011-23125 (MEC).
2 Si a:b :: c:d, entonces a:c :: b:d, según la teoría euclídea de proporciones (*Elementos*, V 16).
3 Por ejemplo Marta Nussbaum, en una entrevista con ocasión de su premio Príncipe de Asturias en CC. Sociales, 2012, opone el rigor de una argumentación como la de Sócrates a "la retórica" –sin mayores precisiones– de los políticos y de los publicitarios. Véanse los extractos recogidos en *Página 12*, 17/mayo/2012 [www.pagina12.com.ar].

días, precisamente, estamos asistiendo a un creciente interés por esa retórica argumentativa debido a su estrecha relación con el discurso público [4].

Pero el título se toma otra licencia retórica más al servirse de una sinécdoque generalizante. En realidad, lo que pretendo elogiar y vindicar es la retórica deliberativa en particular, no de la retórica en general. La retórica deliberativa nace y se desarrolla en la antigua cultura griega de los ss. VI-IV a.n.e. como la forma paradigmática del discurso público. Según es bien sabido, la Atenas de los ss. VI y V a.n.e. alumbra un nuevo régimen democrático con el ascenso del *demos* como un agente socio-político relativamente autónomo, dispuesto a deliberar y decidir colectivamente sobre asuntos de interés común. Entonces aparece la retórica deliberativa como modo de intervenir en el discurso público y, a través de él, en la vida política. Pero su desarrollo pronto concita, de un lado, entusiasmos como los atestiguados por Tucídides, y de otro lado reservas críticas como las de Platón e invectivas satíricas como las de Aristófanes. Así, durante los ss. V-IV, los sofistas y sus competidores van tomando conciencia de esta forma deliberativa de actuación e intervención en los asuntos públicos, y de los problemas y dificultades que genera. Los reparos consisten sustancialmente en negar la legitimidad discursiva y cognitiva de la deliberación pública o la capacidad de una multitud abigarrada para su ejercicio responsable. Por ejemplo, en una asamblea, ¿cómo es posible tomar decisiones racionales, oportunas y efectivas entre tanta gente, gente para colmo indocumentada?

Es instructiva a este respecto la respuesta de Protágoras a una cuestión de este tipo que recoge Platón en los términos del mito de Prometeo-Hermes (*Protágoras* 320d-322). Sócrates pregunta: ¿Por qué se recurre a los expertos en cuestiones técnicas, p. ej. de construcción o navales, pero no cuando se trata del gobierno de la *pólis*? Responde Protágoras: «¿Preferís que lo demuestre [*epideíxo*] relatando un mito o mediante un discurso razonado?» (320c). Con la anuencia de sus interlocutores, Protágoras opta por la demostración a través del mito [5]. Según el mito, Zeus ha encargado a Prometeo la distribución de cualidades entre los seres vivos. Prometeo delega la tarea en un poco previsor Epimeteo que se ve con las manos vacías cuando llega a los humanos. En compensación de sus carencias naturales, Prometeo les entrega el fuego y las técnicas. Pero los hombres vienen a caer entonces en disputas y luchas por intereses enfrentados que hacen imposible la convivencia. Para asegurarse de que al menos sobrevivan, Zeus confía a Hermes un nuevo reparto general e igualitario del sentido de la justicia [*diké*] y de la dignidad o vergüenza [*aidós*]

4 Vid. B. Garsten 2011, "The rhetoric revival in political theory", *Ann. Rev. Polit. Science*, 14: 159-180.

5 Importa reparar en que esta opción es un recurso típicamente retórico a una "historia" o una *narratio*. Así, a la vez que cumple sus funciones expresivas, simbólicas o alegóricas, también pretender servir a los efectos ilustrativos, esto es probatorios y suasorios, de una perspectiva plausible; en suma, quiere ser una respuesta efectiva a la cuestión planteada por Sócrates.

entre los humanos. Pues Zeus repara en que, frente a la anterior distribución desigual de cualidades y recursos, los nuevos dones se han de repartir entre todos los seres humanos porque si solo participaran de ellos unos pocos, como ocurre con las demás artes y competencias especializadas, nunca llegaría a haber ciudades (322c-d).

Aristóteles, por su parte, sienta las bases de una justificación de la deliberación pública en un texto de la *Política* donde expone su concepción del lenguaje o del discurso [*lógos*] como el medio característico de comunicación, de entendimiento y de juicio del ser humano en calidad de animal social.: «El hombre es por naturaleza un animal social [*politicòn zôon*] <…> La razón por la que el hombre es un animal social, más que la abeja o cualquier animal gregario, es evidente: la naturaleza, como solemos decir, nada hace en vano y el hombre es el único animal que tiene *lógos*. La voz es signo de dolor y placer, y por eso la tienen también los demás animales pues su naturaleza les permite sentir dolor y placer e indicárselo unos a otros. Pero el *lógos* es para manifestar lo conveniente y lo perjudicial, lo justo y lo injusto, y es propio y exclusivo del hombre, frente a los demás animales, tener el sentido de lo bueno y lo malo, lo justo y lo injusto, y lo demás, y la comunidad de estas cosas es lo que constituye la ciudad y la casa» (*Política* I.1-2, 1253ª2-18). Por otro lado, «la ciudad es por naturaleza anterior a la casa y a cada uno de nosotros, porque el todo es anterior necesariamente a la parte. <…> Es evidente que la ciudad es por naturaleza anterior al individuo, porque si el individuo separado no se basta a sí mismo, será semejante a las demás partes en relación con el todo, y el que no pueda vivir en sociedad o no necesite nada por su propia suficiencia, no es miembro de la ciudad sino una bestia o un dios» (1253ª19-29). Luego solo las bestias o los dioses serían seres incapacitados o ajenos por naturaleza a la deliberación común. Parece una versión desmitificada del *Protágoras*.

2.

Más allá de esta vindicación, la rehabilitación aristotélica de la deliberación pública discurre a través de su acreditación de la retórica como arte y, en especial, en los términos de la retórica deliberativa. Recordemos sumariamente algunas contribuciones de Aristóteles en este sentido, en particular su concepción de la retórica.

Conviene precisar antes que nada que, según Aristóteles, el propósito de la retórica no consiste en persuadir ni, menos aún, en seducir, sino en considerar teóricamente los medios de generar convicción más pertinentes según sea el caso concreto planteado (*Retórica* I 1.1, 1355^{b}10-12, 25-6). Al ser un arte, consiste en un saber hacer fundado racionalmente en conocimientos verdaderos (*Ética Nicomáquea*, 6.4, 1149ª1-22), y por tanto capaz de explicar su proceder y de justificar sus resultados. La *Retórica* ofrece, en esta línea, dos caracterizaciones expresas.

Por una parte, la retórica es *antístrofa* de la dialéctica (*Retórica* I 1.1, 1354ª1). Las versiones e interpretaciones de este calificativo se han multiplicado con el tiempo. En sustancia denota una contrapartida o correlato de la dialéctica, un arte parejo. Pero la expresión puede tener además una connotación vindicativa: Sócrates, en el *Gorgias* (465d) había calificado la retórica como *antístrofa* del cocinar [6]; así que puede verse en su uso aristotélico una réplica implícita a Platón, gran defensor de los poderes discursivos y noéticos de la dialéctica. Sea como fuere, pueden cifrarse las similitudes o correspondencias de la retórica con la dialéctica en las cuatro siguientes:

(i) No cuenta con un campo de referencia o unos objetos propios, sino que es aplicable a cualquier asunto, (1.1 1354ª2-4, 1.6 1355ᵇ8-9), y por lo tanto no es una ciencia, sino un arte o una disciplina, práctica y productiva por más señas (4.1 1359ᵇ12-16).

(ii) Tiene que ver con el discurso y puede proporcionar pruebas y argumentos razonables y persuasivos (2.2 1355ᵇ35 ss., y 2.3 1356ª20-34).

(iii) Trata con lo plausible y lo verosímil (1.4 1355ª15-18, 27-29), acerca de cuestiones cuya resolución puede ser o considerarse de dos maneras (2.5 1357ª5-6), i.e. demanda o al menos admite debate, y discurre por referencia a lugares comunes (*topicos*) y opiniones más o menos acreditadas (*éndoxa*), i.e. más o menos dignas de crédito en razón de la cantidad y la calidad de la fuente.

(iv) Dispone de cierta capacidad de discriminación entre buenos o genuinos y malos o aparentes usos de los recursos de la disciplina, similar a la competencia de la dialéctica con respecto a las refutaciones efectivas o genuinas y las aparentes o sofísticas.

Pero estas similitudes no ocultan sus diferencias. Una diferencia básica estriba en que el dominio de la retórica es el discurso público. Según esto, bien se puede decir que la retórica es al discurso público lo que la dialéctica es al discurso digamos "particular" o "privado" en este contexto, sea monológico, cuando uno argumenta consigo mismo, o sea dialógico, cuando un agente discursivo discute con otro según el procedimiento normalizado en los *Tópicos*. La referencia al discurso público sitúa la práctica de la retórica, como ya sabemos, en el ejercicio de la política.

Pues bien, esta conexión es la que corresponde a la segunda caracterización formulada por Aristóteles donde la retórica viene a ser «un vástago de la dialéctica y de aquel saber práctico sobre los caracteres [*êthe*] al que es justo denominar política» (2.3 1356ª26-27). En términos modernos

[6] La retórica es a la justicia lo que el cocinar a la medicina (*Gorgias* 465c), o también es respecto del alma lo que el cocinar respecto del cuerpo (465d). En suma, la retórica viene a ser *antístrofa*, correlativa de la cocina (*ibd.*): las dos descansan en la mera experiencia de donde obtienen rutinas o recetas para producir agrado y placer (*Gorgias* 462e), así que una y otra están lejos de constituir artes o ciencias. Platón no bendeciría nuestros cursos universitarios de gastronomía, ni una cátedra como la "Ferrán Adriá" de la Universidad Camilo José Cela.

diríamos entonces que la retórica también es una especie de hijuela de la dialéctica y de una suerte de "política-ética"

Acreditado el estatuto técnico de la retórica, pasemos a considerar sus señas de identidad como disciplina. Para empezar destacan sus objetivos específicos: uno más bien teórico, el otro más bien práctico. El primero ya es conocido por la declaración del propósito de la retórica aristotélica: no consiste precisamente en persuadir, sino en estudiar los medios de persuasión disponibles a través del discurso en cada caso particular. Así cumple además un requisito de su condición de arte: el examen y la explicación de los éxitos de su ejercicio, deliberado o espontáneo (I 1.1, 1354ª10-12). Por otro lado, al atenerse a los medios de persuasión antes que a sus resultados efectivos en la práctica, esta retórica presta especial atención a los aspectos y criterios de eficiencia discursiva y racional interna, un tanto al margen de su eventual eficacia externa. Pero esa eficiencia también guarda relación con la formación de un juicio sobre la cuestión debatida, punto que nos remite al segundo objetivo, el de orden práctico. Consiste en formar un juicio y elegir una opción, por ejemplo un curso de acción, en razón de las pruebas suasorias aducidas.

Las pruebas propias del arte siguen siendo hoy las señas de identidad más famosas de la retórica aristotélica. Dice Aristóteles: «Puesto que la retórica tiene por objeto formar un juicio <…>, es preciso atender no solo a cómo sea demostrativo o digno de crédito el discurso, sino también a cómo ha de presentarse uno mismo [i.e. el hablante o locutor] y a cómo ganarse al que juzga [el alocutario o auditorio]» (II.1 1377b20-24) [7]. Así pues, son estas líneas las que van a marcar las pruebas propias del arte, es decir las que el agente discursivo puede montar mediante su discurso, a diferencia de las pruebas impropias o ajenas con las que puede contar antes o al margen del discurso, como los datos o los testimonios. «Entre las pruebas retóricas, las que pueden obtenerse mediante el discurso son de tres especies: unas residen en el talante [*êthos*] del que habla, otras en poner en cierta disposición [*páthos*] al oyente y las últimas en el discurso [*lógos*] mismo por lo que este demuestra o parece demostrar. Pues bien, <se prueba> por el talante cuando el discurso es dicho de tal forma que hace al orador digno de crédito. <…> Por otro lado, <se prueba mediante la disposición > de los oyentes, cuando estos son inducidos a un estado de ánimo a través del discurso. <…> En fin, los hombres se persuaden por el discurso cuando les mostramos la verdad o lo que parece serlo a partir de lo convincente en cada caso particular» (1.2 1356ª1-19).

Hay dos puntos notables en este planteamiento: 1/ Los tres tipos de pruebas obran a través del discurso, así que hemos de distinguir entre este logos básico, el lenguaje de la comunicación y el

[7] No está de más reparar en la concepción holista o integrada de la formación del juicio que supone esta declaración aristotélica, frente a su tratamiento tradicional como instrucción lógico-epistémica, sin más, o formación primordialmente racional (cf., por ejemplo, la *Lógica o Arte de pensar* de Port Royal).

entendimiento, y el logos argumentativo, el discurso específicamente retórico de las pruebas suasorias o "demostrativas". 2/ Los tres operan conjuntamente, no disyuntivamente, de modo que los procesos de inducción retórica no solo son interactivos, sino que además envuelven a los agentes discursivos en su conjunto, no una facultad en particular (e.g. la racional o la emotiva). El punto 1/ previene contra la búsqueda de una suerte de protológica del discurso racional en el lenguaje comunicativo mismo; el punto 2/ contra las oposiciones al uso que separan el claro rigor del logos de la oscura trama del ethos y pathos en la retórica.

Aristóteles da algunos detalles sobre cada uno de los tipos. Las pruebas argumentativas o correspondientes al logos son el entimema, de estructura deductiva, y el ejemplo, más bien inductivo. Aristóteles sigue siendo fiel a sí mismo y precisa que si bien el ejemplo no es menos convincente, los discursos basados en entimemas están mejor considerados (1356^{b}24). El entimema, en particular, es «el cuerpo de la persuasión» (1354ª15) y se dirige no solo al entendimiento [8], sino al hombre completo, dado el carácter integrado u holista tanto de la motivación del discurso retórico como de su eficiencia práctica en orden a la formación del juicio. El entimema representa, por lo demás, una de las muestras más elocuentes de la cooperación entre los distintos tipos de pruebas, pues su efectividad suasoria descansa en la comprensión y la complicidad del interlocutor o del oyente [9].

A su vez, «el talante constituye por así decir el más firme medio de persuasión» (1356ª13), pero a condición de que obre a través del discurso. Según Aristóteles, se persuade «por el êthos cuando el discurso es dicho de tal forma que hace al orador digno de crédito» (1356ª5-6). Hay tres causas o motivos concurrentes de que un agente discursivo sea fiable o digno de crédito: (i) la prudencia o inteligencia práctica (*phrónesis*), (ii) la virtud (*areté*) y (iii) la buena disposición o franqueza (*eúnoia*). La mera manifestación de inteligencia práctica, no acompañada de signos de virtud ni de buena disposición, induciría a dudar sobre las intenciones del orador; mientras que las señales de prudencia y de virtud, sin muestras de franqueza, inducirían a dudar sobre su grado de implicación o compromiso con lo que propone.

Y, en fin, el poder de convicción del entimema y el reconocimiento del talante fiable del emisor requieren a su vez la buena disposición, el oportuno estado de ánimo y la complicidad del receptor. Cabe decir, en general, que el orador persuade a los oyentes «cuando estos son movidos a un estado de ánimo por medio del discurso», 1356ª13-14. Ahora bien, el pathos aristotélico no deja de tener una dimensión cognitiva pues supone considerar el estado del sujeto, el objeto de la

[8] «El intelecto solo <…> no mueve nada» (*Ética Nicomáquea* 1139b35-36).

[9] Sobre la recuperación actual de este y otros aspectos olvidados por la tradición empeñada en tratar el entimema como un razonamiento defectivo o un silogismo deficiente, vid. mi (2004), "Entimemas", *Doxa*, 27: 283-315.

emoción producida y el motivo de esta inducción (en el libro II de la *Retórica* 1.3, 1378ª23-24, puede verse un examen de la ira según estos tres parámetros). De ahí que sea preferible entenderlo no como pasión, en la medida en que las pasiones incluyen pulsiones no cognitivas, sino en el sentido de emoción o disposición de ánimo inducida a través del discurso.

Por último, de los géneros retóricos, el directamente referido al discurso público es el deliberativo. Su propósito es persuadir de lo conveniente y disuadir de lo perjudicial en determinados asuntos de interés común para el conjunto de la ciudad. Pues, según advierte Aristóteles, no deliberamos sobre cualquier cosa. Para empezar, deliberamos sobre lo que depende de nosotros y nos concierne. Así que no deliberamos sobre lo necesario, lo inmutable o lo eterno, porque todo esto sobrepasa nuestra capacidad de actuación. Tampoco deliberamos sobre el puro azar porque también está fuera de nuestro control y ocurre al margen de nuestras posibilidades de intervención. Y, en fin, una comunidad de ciudadanos tampoco delibera sobre lo que no le concierne como, por ejemplo, una constitución política ajena.

3.

En este punto de la deliberación pública es donde podemos sentir a Aristóteles más lejos de nosotros. Solo un par de indicaciones al respecto. 1/ Una tiene que ver con el discurso deliberativo mismo: Aristóteles se hace una idea limitada del razonamiento práctico pertinente, ahormado en la línea instrumentalista medios-fines, y no considera ingredientes que hoy nos parecen relevantes, como los valores o los compromisos, amén de ignorar el papel de la ponderación de alternativas y la discusión acerca de propuestas. En este sentido, su retórica se resiente de deficiencias no solo sustantivas, e.g. en el terreno de la ética y la política de los fines, sino analíticas, es decir lógicas y dialécticas. 2/ La otra señal de lejanía es la distancia que existe entre su micro-deliberación cara a cara, mediada por el êthos de la propia reputación y por la vergüenza de hallarse ante los ojos de los otros –según el proverbio que cita en la *Retórica*, 1384ª34–, y nuestra macro-deliberación mediante interfaces de dispositivos en red, donde se han sustituido voces y miradas por direcciones IP en una comunicación, digamos, “desvergonzada”

Claro está que en estas, como en otras cuestiones relacionadas con la retórica deliberativa y con el discurso público, nuestra ventaja sobre Aristóteles no reside en haberlas resuelto, sino en haber llegado a conocerlas y plantearlas.

PHILOSOPHY OF LANGUAGE, PHILOSOPHY OF MIND, EPISTEMOLOGY

FILOSOFÍA DEL LENGUAJE, FILOSOFÍA DE LA MENTE, EPISTEMOLOGÍA

PUNTOS DE VISTA

Organizadores:
JUAN J. COLOMINA y MANUEL LIZ
The University of Texas at Austin y Universidad de La Laguna
Colomina-Alminana_Juan@austin.utexas.edu y manuliz@ull.es

El tema de la mesa redonda es la noción de puntos de vista. Esta noción desempeña un papel teórico crucial en numerosos campos, tanto dentro como fuera de la filosofía. Y también desempeña un importante papel en nuestras concepciones ordinarias del mundo y de nosotros mismos. Sorprende, por ello, que apenas haya recibido análisis específicos. Mientras que la noción de puntos de vista es abundantemente usada en prácticamente cualquier contexto, la literatura filosófica que aborda directamente un análisis de dicha noción es sumamente escasa.

Nuestra mesa redonda propone una reflexión sobre este hecho sin duda singular. Y los participantes invitados en ella sugerirán algunos enfoques que merecería la pena tener en cuenta, y acaso explorar con mayor detalle.

Introducción

En la filosofía contemporánea, el uso de la noción de puntos de vista ha sido sumamente abundante. Este uso tiene lugar en muy diversas áreas y siempre vinculándose a problemas sumamente debatidos. Algunos de ellos son el del relativismo y la inconmensurabilidad, en epistemología y filosofía de la ciencia; el de la subjetividad, en filosofía de la mente; el del contraste entre internismo y externismo respecto al significado, en filosofía del lenguaje; o el del contenido objetivo de los juicios morales, en ética.

Fuera de la filosofía, la noción de puntos de vista (o perspectivas) también es relevante en múltiples contextos, pudiendo llegar a tener una significación muy amplia. La expresión "punto de vista" aparece involucrada crucialmente en infinidad de descripciones que se refieren a nuestra vida personal y al mundo social, haciéndose particularmente visible cuando va acompañada de posiciones de tipo relativista. Y también existen importantes usos de esa expresión dentro de los ámbitos científicos y técnicos.

Puede decirse que la noción de puntos de vista ha ocupado un lugar central en muchas discusiones filosóficas actuales, que también desempeña un papel relevante dentro de la ciencia y de la tecnología, y que de un modo u otro está profundamente arraigada en nuestras concepciones del mundo y de nosotros mismos. Frente a esto, llama la atención que la noción de puntos de vista apenas haya recibido hasta el momento tratamientos directos y sistemáticos. Tenemos algunos antecedentes en las filosofías relativistas clásicas y modernas de Heráclito, Protágoras, Kant o Nietzsche. También contamos con los perspectivismos de autores como Leibniz, Nietzsche, James y nuestro Ortega (al que tal vez podríamos denominar "el primer perspectivista explícito"). Y también disponemos de algunos análisis contemporáneos de interés indudable para un análisis de la noción de puntos de vista. Podríamos referirnos aquí a una larga lista de autores que inevitablemente tendría que incluir a Kuhn, Davidson, Putnam, Goodman, Rorty, Searle, Nagel, Williams, o Cavell. Y sobre todo, debemos mencionar aquí los trabajos, excepcionalmente más específicos, de Jon Moline, Christopher Peacocke, Antii Hautamäki y Adrian Moore. Estos últimos autores sí han abordado directamente la tarea de analizar la naturaleza de los puntos de vista. Con todo, el tema ha sido tratado de modo muy tangencial a otros temas y problemas.

2. Objetivo

Nuestro objetivo es promover una reflexión sobre la noción de puntos de vista; sobre su naturaleza, su estructura, su dinámica y su modo propio de existencia. Esta reflexión se situaría entre los campos de la lógica, la filosofía de la mente, la filosofía del lenguaje, la epistemología y la ontología.

En este sentido, lo primero que debe señalarse es una serie de ambigüedades que rodean la noción de puntos de vista. Hay dos ambigüedades particularmente importantes. La primera tiene que ver con el carácter subjetivo u objetivo de los puntos de vista. Por un lado, muchas veces los puntos de vista se vinculan a ciertas clases de sujetos, a sus capacidades y actitudes psicológicas, tanto conceptuales como no conceptuales. Por otro lado, los puntos de vista también se asocian a ciertos emplazamientos, o localizaciones, y posibles accesos a ciertos tipos de objetos o de verdades, desde esos emplazamientos. En el primer caso, se enfatizan los aspectos subjetivos de los puntos de vista. En el segundo caso, se enfatizan sus aspectos objetivos. Esta ambigüedad es muy importante. Y está presente en todos los usos de la noción de puntos de vista; tanto en contextos filosóficos, como en contextos científicos y ordinarios.

La segunda ambigüedad se refiere al poder constitutivo, o configurador, de los puntos de vista sobre cosas como el contenido, la verdad, el conocimiento o la acción. A veces, ese poder

constitutivo se considera máximo. Esto ocurre cuando, por ejemplo, los puntos de vista se vinculan a los lenguajes o esquemas conceptuales. Y cuando se pretenden consecuencias fuertemente relativistas o limitadoras. Otras veces, el poder constitutivo es mínimo. O sólo afecta, por ejemplo, a la selección de "aspectos". Esto es lo que ocurre con los perspectivismos que quieren marcar diferencias respecto al relativismo. Esta ambigüedad no es menos importante que la primera. Pero se trata de dos ambigüedades diferentes.

Cada una de las anteriores ambigüedades hace inclinar el análisis de los puntos de vista hacia determinadas posiciones. En relación con la primera ambigüedad, pueden distinguirse dos orientaciones básicas. Una de ellas adoptaría como modelo las actitudes proposicionales, ampliando ese modelo para incluir todos los aspectos no-conceptuales (perceptivos, agentivos, evaluativos y normativos) de los puntos de vista. La otra orientación adoptaría como modelo las nociones mismas de emplazamiento y acceso. En relación con la segunda ambigüedad, nuestra orientación podría variar dentro de un amplio espectro de posiciones perspectivistas cuyos extremos serían los de un relativismo máximo y alguna versión de la "concepción absoluta" del mundo y de nosotros mismos.

Más allá de la discusión de estos temas generales, existe un gran conjunto de temas que debería abordar un análisis en profundidad de la noción de puntos de vista. Cabría mencionar los siguientes: puntos de vista individuales, colectivos e intersubjetivos; puntos de vista personales; contenidos conceptuales y no-conceptuales de un punto de vista; puntos de vista reflexivos; puntos de vista evaluativos y normativos; puntos de vista agentivos; puntos de vista temporales; puntos de vista trascendentales; puntos de vista científicos y filosóficos, etc.

Uno de los desafíos más importantes del análisis de los puntos de vista es el de sopesar la posibilidad de articular una concepción de los mismos capaz de evitar el subjetivismo, el relativismo, el trascendentalismo, la explosión de hechos dependientes de una perspectiva y el reduccionismo. Acaso, la pregunta más radical al respecto sea la de cómo es posible la existencia de puntos de vista en un mundo físico.

Varios desarrollos de algunos de estos temas han sido tratados en dos proyectos de investigación llevados a cabo desde la Universidad de La Laguna en estos últimos años. Dichos proyectos son:

Puntos de vista. Una investigación filosófica (Referencia FFI 2008-01205)

Puntos de vista y estructuras temporales (Referencia FFI 2011-24549)

Las reflexiones y análisis que se ofrecen en la mesa son, por tanto, resultado de unas líneas de investigación actualmente en marcha. Respecto a ellas, es destacable el hecho de que al explorar la geografía conceptual vinculada a la noción de puntos de vista, haya sido necesario profundizar en

los aspectos temporales de los mismos. Las estructuras temporales que podrían ser relevantes en la dinámica de los puntos de vista, sobre todo en puntos de vista con cierto grado de complejidad, pueden acabar siendo esenciales a la naturaleza misma de los puntos de vista, y a su modo peculiar de existencia.

Resumen de las participaciones

¿SON SUBJETIVOS LOS PUNTOS DE VISTA?

MANUEL LIZ
Universidad de La Laguna

ABSTRACT: The paper argues against the thesis that points of view, or perspectives, are necessarily subjective entities. On the one hand, there are many uses of the notion of points of view which cannot be interpreted in a mere subjective sense. On the other hand, the available theoretical options in order to analyse the structure and nature of points of view do not support either the thesis that points of view are subjective. Points of view have become the paradigm of subjectivity because they are supposed to entail some kind of Protagorean relativism and because they are supposed to be reducible to our subjectivity. We reject both arguments.
KEY WORDS: Points of view, perspectives, subjectivism, relativism, mind.

* * *

La respuesta a esta pregunta parece obvia. La pregunta misma tiene un aire retórico que también parece obvio. Los puntos de vista "han" de ser subjetivos. Más aún, los puntos de vista "han" de ser el paradigma mismo de la subjetividad, el "estándar" de lo subjetivo. En mi intervención, quiero poner en cuestión estas presuntas obviedades.

Mi estrategia se desarrolla en dos frentes. El primero de ellos lo constituye el uso de expresiones como "desde mi punto de vista", y pensamientos asociados. En segundo frente es el de los enfoques teóricos que cabe adoptar respecto a la naturaleza y estructura de los puntos de vista.

En el primer frente, comenzaré destacando la gran variedad de usos que tienen expresiones del tipo "desde mi punto de vista" ("desde nuestro punto de vista", "desde mi/nuestra/etc. perspectiva", "a mi/nuestro/etc. modo de ver", "en mi/nuestra/etc. opinión", etc.). Entre todos estos usos, hay muchos que podrían denominarse "usos protagoreanos". Y tales usos suscitan, ciertamente, una fuerte relativización subjetiva de los contenidos a los que se aplican esas expresiones. Sin embargo, hay otros muchos usos que no van acompañados de esa clase de relativización. Compárese, por ejemplo, (1) "El café está muy amargo, desde mi punto de vista" con (2) "La decisión del tribunal no fue la correcta, desde mi punto de vista". La intención que anima 2

no es precisamente relativizadora, ¡sino justamente lo contrario!. Sorprendentemente, los usos tipo 2 son de hecho muy abundantes. Y pasarlos por alto hace que se considere que todo enunciado con una forma del tipo "p, desde mi punto de vista" implique una relativización subjetiva. Pero esto, simplemente, no es cierto.

En el segundo frente, distinguiré dos opciones teóricas muy generales a la hora de abordar el análisis de la naturaleza y estructura de los puntos de vista. Una de estas opciones adoptaría el modelo de las actitudes proposicionales. Un punto de vista sería aquí una estructura compleja incluyendo sujetos, actitudes de estos sujetos hacia una serie de contenidos, contenidos tanto no conceptuales como conceptuales, y determinadas condiciones de posesión del punto de vista. La otra opción se centraría en las nociones de emplazamiento y acceso. A grandes rasgos, un punto de vista sería un mundo desde el que puede accederse a otros mundos. Cabe mencionar autores muy relevantes y líneas específicas de desarrollo de cada una de las anteriores opciones teóricas.

¿Son subjetivos los puntos de vista en alguno de estos dos enfoques teóricos? El segundo enfoque está claramente descomprometido con el subjetivismo de los puntos de vista. La relación de accesibilidad que involucra puede ser sumamente objetiva. Desde ciertos emplazamientos se accede a ciertas cosas, y esto puede ser un rasgo plenamente objetivo de la realidad.

El primer enfoque, parece tener un compromiso mayor con el carácter subjetivo de los puntos de vista. Algunos autores recientes se han apoyado en esta idea (por ejemplo, T. Nagel, o K. Farkas con sus propuestas neo-cartesianas). Sin embargo, hay algo confuso en todo esto. Debería distinguirse entre, por un lado, lo interno y lo externo a un punto de vista y, por otro lado, lo interno y lo externo a los sujetos (a sus mentes, etc.) que pueden llegar a adoptar dichos puntos de vista. No se trata de la misma distinción. Cabe argumentar que los contenidos de un punto de vista, tanto no-conceptuales como conceptuales, o las propias condiciones de posesión del punto de vista, son algo interno al punto de vista, sin por ello ser algo interno a los sujetos capaces de adoptar tal punto de vista. Esta distinción es sumamente importante. Y permite evitar la ilusión de que el primer enfoque sobre los puntos de vista, el basado en el modelo de las actitudes proposicionales, sí presta apoyo a una interpretación subjetivista de los puntos de vista.

No apoyan el subjetivismo ni los usos de expresiones como "desde mi punto de vista" ni, tampoco, los enfoques teóricos disponibles sobre los puntos de vista. ¿Por qué, entonces, está tan extendida la idea de que los puntos de vista son esencial e irreduciblemente subjetivos, que son el paradigma, el estándar, de lo subjetivo?

Hay dos factores ideológicos que me parecen cruciales. Uno de ellos es la estrecha vinculación de la noción de punto de vista con el relativismo de Protágoras. El otro es el frecuente recurso

a la noción de punto de vista cuando se trata de identificar fenómenos mentales. Pero el relativismo protagoreano es rechazable. Ni siquiera es la variedad más sugerente de relativismo (en comparación, por ejemplo, con un relativismo inspirado en Heráclito). Y la apelación a los puntos de vista en la identificación de fenómenos mentales sólo podría hacer de la mente algo esencial e irreducibleme nte subjetivo si los propios puntos de vista necesariamente fueran algo esencial e irreducible-mente subjetivo. Y no es el caso.[1]

[1] Este trabajo ha sido llevado a cabo en el seno de los proyectos de investigación FFI 2008-01205 y FFI 2011-24549.

PUNTOS DE VISTA COMO ACCESO

JUAN J. COLOMINA y FERNANDO CHARRO
The University of Texas at Austin

ABSTRACT: This paper identifies points of views by their role. Here, they can solely be understood as ways of accessing the world that bring some kind of perspective about it.

Having this in mind, we would like to propose a notation that explains how to understand such access as a sort of conceptualization model, independently of whether the precise point of view under consideration is impersonal or non-impersonal, its kind of content, and its subjective or objective character. First, we will analyze what kind of structure the world is assumed to posses and how the access to it is possible. Then, we will develop some notions that explain points of view as qualitative dimensions by means of which it is possible to valuate objects and states of the world.

KEY WORDS: Points of view, qualitative dimension, filter function, perspective, states of affairs.

* * *

Un modo de analizar la noción de puntos de vista involucra las nociones de localización y acceso. Desde esta posición, la estructura interna de los puntos de vista no es especialmente relevante. Aquello que se enfatiza son las propiedades de los puntos de vista relacionados con las funciones que éstos desempeñan. Esto es, con su rol. Vistos así, los puntos de vista sólo pueden comprenderse como modos de acceso al mundo desde algún tipo de perspectiva sobre el mismo. Lo que se pretende en esta presentación es aportar algunas nociones que expliquen los puntos de vista como dimensiones cualitativas desde las que evaluar los objetos y los estados del mundo.

Partiremos de la asunción de que el mundo físico tiene una estructura de determinables y determinados. Los puntos de vista, así, son capaces de considerar diversos objetos del mundo. Todos ellos ejemplificarían la relación existente entre determinados (propiedades particulares o cualidades de segundo nivel) que pertenecen a un cierto determinable (un término universal entendido como un conjunto de cualidades o propiedades de primer orden). Según la mencionada estructura, supondremos que el mundo tiene de un modo objetivo propiedades de primer orden que pueden ser refinadas en diferentes determinados.

Los puntos de vista, entendidos aquí como entidades pertenecientes al mundo físico, heredan esta estructura. Visto de este modo, un particular punto de vista estructura el mundo a partir de un número concreto de clases que se instancian en las características reales del mundo físico (PW, del

inglés physical world). Incluso en el supuesto de la existencia de un punto de vista con un infinito numerable de determinables, se asume que el número de determinables disponibles al punto de vista será finito.

Al analizar desde esta perspectiva la noción de punto de vista llegamos a una aparente paradoja. Los determinables existen por sí mismos como características actuales de los objetos del mundo físico, y en este sentido existen de una forma independiente de los puntos de vista que tienen acceso a ellos. Al mismo tiempo, sin embargo, los puntos de vista crean tales determinables a partir del acto de acceso al objeto. En este sentido, el conjunto de determinables del mundo a los que tiene acceso los puntos de vista serían un modelo del mundo físico. De este modo, los determinables tendrían un estatuto dual. Por una parte, los determinables son características físicas de los objetos. Pero, por otra parte, los determinables son el modo desde el que los puntos de vista tienen acceso al mundo.

Con el objetivo de aunar ambos aspectos, introducimos la noción de mundo filtrado. Dado un particular punto de vista, digamos O, consideraremos una función de filtro f_o. El input de esta función de filtro será siempre el mundo físico, mientras que su output será el conjunto de determinables a los que O tiene acceso. Podemos llamarlos determinables disponibles, y $f_o(PW)$ será la denominación que daremos al mundo filtrado. La principal ventaja de este tipo de notación es que permite formular de una manera intrínseca nuestro argumento sólo en términos de determinables que son accesible al punto de vista O, incluso aceptando que parte del mundo físico quede fuera del rango de ese particular punto de vista.

Sin embargo, existe una dificultad en esta descripción sobre aquello que es un determinable. Deberemos suponer que ciertos determinables son multidimensionales. En este sentido, debemos entender el acceso a ciertos objetos desde una noción de función de estado. Así, la función de estado S instancia una entidad del mundo dentro de un determinado espacio de estado, entendido éste como un producto cartesiano de todo posible determinable que se percibe desde el punto de vista en cuestión. Si esto es así, la función de estado filtraría las entidades del mundo físico disponibles a los puntos de vista y produciría los vectores necesarios que darían lugar a los determinados potencialmente accesibles.

Para ilustrar esta línea de razonamiento presentamos el siguiente ejemplo. Imagina que el mundo filtrado del punto de vista O, esto es $F_o(PW)$, consiste de las propiedades 'años (de edad)' (A), 'peso (en kilogramos)' (W) y el espacio de estado resultante del producto cartesiano AxW. Supón que el mundo al que O tiene acceso consiste tan sólo de dos individuos: los seres humanos Ulises y Penélope. Podemos presentar la información a la que O tiene acceso del siguiente modo:

S(Ulises) = (40 A, 180 W)

y

S(Penélope) = (20 A, 125 W).

De este modo, se hace evidente cómo desde el punto de vista O tanto Ulises como Penélope son objetos multidimensionales.

Para clarificar estas nociones, introduciremos un tercer objeto: el ser humano Jasón. Supón ahora que Jasón tiene exactamente las mismas características que Ulises, según este universo simplificado al que tiene acceso O.

S(Jasón) = (40 A, 180 W).

En este caso, Jasón y Ulises son indistinguibles desde el punto de vista O. Pero al mismo tiempo, ambos objetos son perfectamente identificables como objetos diferentes del mundo filtrado, ya que el espacio de estado consiste tan sólo de las mencionadas características A y W. Parece que debemos admitir la existencia de una tercera característica implícita a partir del hecho de que Jasón y Ulises aparezcan como objetos diferentes desde el mundo filtrado de O.

Es más, podríamos expresar algunas de las características del mundo filtrado de O en términos de las otras características. Esto puede entenderse mejor si repensamos nuevamente nuestro anterior ejemplo, en donde tanto Jasón como Ulises se perciben como objetos diferentes desde el mundo filtrado de O. Así,

S(Jasón) = (40 A, 180 W, Jasón R)

y

S(Ulises) = (40 A, 180 W, Ulises R).

En este ejemplo estamos explicitando la tercera característica que insinuamos como interna y a la que denotamos con R (por 'aquello que resta'). Esta característica equivale al conjunto restante constitutivo de Ulises y de Jasón respectivamente y que no son ni A ni W. Formulado de este modo, en realidad podemos describir la propiedad R en términos de {A,W}. Sin embargo, nuestra noción R tiene la ventaja de una mayor flexibilidad rotacional, dado que puede ser refinada tanto como nos sea necesario. En otras palabras, podemos separar R en determinables más refinados y podemos concebir R como un vector (en principio, R podría tener un número cualquier de coordenadas que van desde el 0 hasta el número total de determinables a los que el punto de vista O tiene acceso).

De este modo, los determinables se identifican a partir de sus determinados en tanto que son percibidos en dos o más objetos del mundo. En principio, tales determinados se pueden describir desde un objeto del mundo como una oración del tipo 'aquella característica de dichos objetos que

no es ni esto ni aquello'. Este proceso se hace mucho más complejo cuando el número de determinables a los que O tiene acceso se incrementa, y esto es algo que repercutirá a la hora de definir los determinados (en virtud de un principio de economía ontológica).

Asumamos ahora que en el mundo al que O tiene acceso existe otro objeto, esta vez de una clase diferente: el perro Argos. Argos se percibe desde el particular mundo filtrado de O como

S(Argos) = (12 A, 40 W, Argos R).

Recuerda que el mundo al que punto de vista O tiene acceso es excesivamente limitado en nuestro ejemplo, donde sólo existen cuatro objetos y se pueden identificar tres propiedades. El punto de vista O será capaz de percibir las características A, W y R y ello será suficiente para individualizarlos. Sin embargo, como ya hemos mencionado anteriormente, puede parecer natural para O refinar cada vez más la propiedad R y diferenciar un nuevo determinable todavía más preciso, digamos K (por la palabra inglesa para clase: 'kind'). Desde el mundo de objetos a los que el punto de vista O tiene acceso en nuestro ejemplo, el determinable K tendría los determinados mutuamente exclusivos 'humano' y 'perro'. De este modo, el total del mundo al que el punto de vista O tiene acceso sería definido como

S(Ulises) = (40 A, 180 W, humano K, Ulises R)
S(Penélope) = (20 A, 125 W, humano K, Penélope R)
S(Jasón) = (40 A, 180 W, humano K, Jasón R)
S(Argos) = (12 A, 40 W, perro K, Argos R).

Este proceso de refinamiento puede extenderse todavía mucho más. Pero necesariamente deberá tener un número finito de pasos en tanto que se asume que el número de determinables a los que el punto de vista O tiene acceso (el número de objetos y características del mundo filtrado de O) es finito.

Siguiendo el espíritu de nuestro ejemplo, introducimos una nueva noción: la noción de sistema de generadores. A partir de la misma, diremos que un conjunto S de objetos del mundo filtrado de O, $F_o(PW)$, es un sistema de generadores de ese mundo filtrado en tanto que S abarca el mundo filtrado de O en términos de algún tipo de valores (o determinados) de un cierto tipo de propiedades (o determinables). Date cuenta de que visto de este modo el mismo mundo filtrado se define como un sistema de generadores. Sin embargo, podría suceder que los elementos de $F_o(PW)$ podrían ex-

presarse en términos de un subconjunto más pequeño de S, obligando así a reducir el sistema de generadores mismo.[2]

LA LÓGICA AFECTIVA DE LOS PUNTOS DE VISTA

DAVID P. CHICO
Universidad de Zaragoza

ABSTRACT: Generally speaking, the notion of point of view has been underestimated in philosophy. Nonetheless, it can help to shred new light on many contemporary philosophical debates. Here we want to put this claim to the test in relation to the quarrel between realism and anti-realism and the question of how does our cognitive activity constitute the reality we perceive. First we want to reconsider the idea according to which it is our responsiveness to the salience of the world (of many of its features) what articulates our interaction with it; and secondly we would like to show that this responsiveness follows what might be called an affective logic.
KEY WORDS: Points of view, realism, anti-realism, affective-logic.

* * *

En una visión del mundo como la que en la Modernidad se va configurando a partir de la filosofía de Descartes y de los empiristas británicos, la subjetividad de nuestros puntos de vista es vista como un impedimento formidable a la hora de acceder epistémicamente a una realidad que existe con independencia de las representaciones que nos podamos hacer de la misma. En esta visión filosófica del mundo se dan cita el realismo metafísico y una teoría causal de la percepción. La historia de la filosofía moderna es la suma de los distintos intentos que trataron de tender puentes sobre el abismo que se abre entre el sujeto cognoscente y el mundo como un objeto más (acaso el más grande) de conocimiento. En el siglo XX nos encontramos con ecos de esta problemática, por ejemplo, en la disputa entre realistas y antirrealistas, y en la argumentación a favor de una concepción absoluta de la realidad por parte de Bernard Williams y los comentarios de Thomas Nagel en torno a la idea de la visión (el punto de vista) desde ninguna parte. Williams justifica la necesidad de disponer de una concepción absoluta de la realidad para superar la subjetividad de nuestros puntos de vista y lograr así conocimiento de la realidad. Nagel contesta a Williams con una concepción que no confunde racionalismo con subjetivismo. En todo caso, ambos autores participan de una misma o parecida concepción de la realidad en la que el mundo es un objeto y la relación básica que mantenemos con él es una de conocimiento.

[2] Auspiciado por los proyectos "Puntos de vista. Una investigación filosófica" (FFI 2008-01205) y "Puntos de vista y estructuras temporales" (FFI 2011-24549).

Aquí nos proponemos revisitar los debates anteriores a partir de intentos como los de Hilary Putnam o de John McDowell de encontrar una vía alternativa al realismo metafísico que no incurra en el antirrealismo. Putnam, por ejemplo, ha señalado que el subjetivismo se alimenta de la dificultad de responder a cómo es posible que unas criaturas dotadas de las capacidades contingentes de una determinada especie biológica, podamos tener acceso a métodos universalmente válidos de pensamiento objetivo. En otras palabras, el realismo metafísico y el antirrealismo serían dos caras de una misma moneda.

Daremos por bueno el diagnóstico de Putnam, e intentaremos ir más allá situando la noción de punto de vista en el centro del debate y proponiendo que nuestra relación más básica con el mundo se constituye a partir de una cierta lógica afectiva. Nos centraremos en la propuesta de McDowell según la cual mantenemos el mundo a la vista (se trata la piedra de toque de nuestras representaciones, el tribunal ante el que responden nuestras pesquisas y pretensiones de conocimiento) pero a una cierta distancia (la suficiente como para que la normatividad que rige en nuestra propia concepción del mundo tenga que ver con la coherencia de ésta). En otras palabras, nuestras representaciones y pretensiones de conocimiento verdaderas son acerca del mundo, pero ninguna justificación que podamos dar de las mismas garantiza su verdad. Seguiríamos manteniendo la diferencia entre cómo es el mundo y cómo nos parece que es: nuestros puntos de vista no tienen acceso a la realidad objetiva e independiente de nuestras representaciones de la misma. McDowell ha defendido recientemente que nuestra percepción es un medio, una capacidad, para obtener conocimiento; o, también, que nuestras capacidades perceptivas son modos de receptividad de los rasgos de nuestro entorno. Ahora bien, nos parece que esta receptividad o la manera en que nuestra concepción del mundo responde ante el mundo, no se explica satisfactoriamente con una mera relación causal o sugiriendo que la manera como es el mundo es una norma para la verdad y la justificación. Con la noción de "lógica afectiva", que tomamos de Kathleen Lennon, queremos ver si es posible completar la propuesta de McDowell. Lo que tenemos en mente se explica igualmente recordando las intuiciones de Stanley Cavell acerca de que el mundo es un lugar significativo que demanda nuestra atención y exige nuestras respuestas.[3]

REFERENCIAS

Dummett, M. 1993, "Realism and Anti-Realism", en *The Seas of Language*, Oxford, Clarendon Press, 462-478.

[3] Proyectos: "Puntos de vista. Una investigación filosófica" (FFI 2008-01205) y "Puntos de vista y estructuras temporales" (FFI 2011-24549).

Lear, J. 1982, "Leaving the World Alone", *The Journal of Philosophy* 79(7): 382-403.
Lennon, K. 2010, "Re-enchanting the World: The Role of Imagination in Perception", *Philosophy* 85, 375-389.
McDowell, J. 1981, "Anti-realism and the epistemology of Understanding", en H. Parret y J. Bouveresse (eds.), *Meaning and Understanding*, Berlin, de Gruyter.
---. 1984, "Wittgenstein on Following a Rule", *Synthese* 58, 325-363.
---. 2011, *Perception as a capacity for knowledge*, Milwake, Marquete University Press.
Putnam, H. 1994, "Sense, Nonsense and the Senses. An Inquiry into the Powers of the Human Mind", *Journal of Philosophy* 91, 445-517.
Wright, C. 1986, "Rule Following, Meaning and Constructivism", en C. Travis (ed.), *Meaning and Interpretation*, Oxford, Blackwell.
Zalabardo, J. L. 2000, "Realism Detranscendentalized", *European Journal of Philosophy* 8, 63-88.

ATRIBUCIÓN TEMPORAL DE UN PUNTO DE VISTA

MARGARITA VÁZQUEZ
Universidad de La Laguna

ABSTRACT:
In this paper, it is analyzed the fact that logical analysis of time has important consequences in philosophy, especially for the study of the notion of point of view. It is also studied the new perspective that hybrid logic could give to this topic.
KEY WORDS: Point of view, Time, Temporal logic, Ockhamist logic, Hybrid logic

* * *

El análisis lógico del tiempo, y de manera especial la perspectiva que da la lógica híbrida sobre el análisis formal del tiempo, tiene importantes consecuencias en filosofía. Estas consecuencias son especialmente manifiestas para el estudio de la noción de punto de vista.

Hay dos maneras clásicas muy conocidas de entender el tiempo. Desde una perspectiva externa, el tiempo se describe como una sucesión de momentos o instantes, que pueden ser enumerados.

Frente a esta perspectiva externa, hay una perspectiva interna, de acuerdo con la cual necesitamos tener en cuenta el factor humano para entender completamente el tiempo. Tenemos que introducir los tiempos verbales y considerar el tiempo en términos del paso del tiempo. La lógica temporal, tal como fue desarrollada por Arthur Prior, pretende capturar esta segunda perspectiva interna a través del uso de una lógica similar a la modal. Él creía que, de esta manera, era posible capturar el modo en el que el lenguaje se refiere al tiempo. Pero esto no es posible debido a la incapacidad de la lógica temporal para referirse a momentos particulares del tiempo. Una manera de superar esta limitación es *hibridizando* la lógica temporal.

Cuando analizamos los puntos de vista y su estructura, parece claro que hay que tener en cuenta el momento *t* en el que un punto de vista se atribuye a un sujeto. Esta atribución podría consistir en la adscripción a un sujeto de un cierto conjunto de proposiciones, o de objetos y propiedades, en un momento *t* bajo ciertas condiciones.

Una lógica temporal bidimensional, como la llamada por Prior "Ockhamista", que combina una dimensión modal para los futuros posibles y una temporal para el flujo del tiempo, ofrece una visión muy clara del tiempo, visto como un flujo indeterminista. En esta lógica, tenemos historias, gracias a la dimensión temporal, en las que la relación entre los momentos del tiempo es lineal-irreflexiva, y la evaluación es en dos índices, el momento y la historia. Estas historias diferentes pueden ser vistas como distintos escenarios, lo que proporciona una ventaja indudable a la hora de aplicarlo al análisis de los puntos de vista. Pero tiene también importantes limitaciones. La limitación más importante para nuestro análisis es que no permite hacer referencia a momentos particulares del tiempo. Es decir, no se puede establecer el momento exacto *t* en el que una proposición (o un conjunto de proposiciones) atribuida a un sujeto es verdadera.

Esta limitación no es exclusiva de la lógica ockhamista, sino que es un problema común a la mayoría de las lógicas temporales unidimensionales o bidimensionales. Es una limitación grave, dado que la semántica del lenguaje natural hace un uso constante de la referencia temporal y las lógicas temporales no son capaces de dar cuenta de ella. Este es uno de los aspectos de la lógica temporal que la lógica híbrida parece solucionar, extendiendo la lógica de base con un mecanismo para referirse a los momentos del tiempo. Además, con la extensión híbrida, se mantiene la perspectiva interna del tiempo, dado que las fórmulas se evalúan dentro de los modelos, en un momento del tiempo concreto. Patrick Blackburn (Blackburn (2006)) hace un símil de una fórmula modal con una criatura (o un autómata) situada dentro de una estructura en un punto y forzada a explorar haciendo transiciones a puntos accesibles. Podría hacerse un símil semejante con respecto a las fórmulas temporales situadas en un momento *t* y trasladándose a momentos anteriores y posteriores.

La ayuda que proporciona la lógica híbrida se ve clara con un ejemplo. En un enunciado como "Ulises había visto el libro", sin ayuda de la lógica híbrida no somos capaces de hacer referencia al momento exacto de tiempo *t* en el que es verdad que Ulises lo vio. Sí podríamos decir que si la proposición "Ulises había visto el libro" es verdadera en *t*, entonces hay un momento anterior, *s*, donde es verdadera que "Ulises vio el libro" y otro anterior a este, *r*, donde "Ulises ve el libro". Si tenemos otra proposición verdadera en *t*, por ejemplo "Ulises había pensado en un caballo", no podemos señalar si el momento donde es verdad que "Ulises pensó en un caballo" coincide con *s* o no. La lógica híbrida nos permite, justamente, por medio de los nominales, poner una marca a ese momento *s* y ver si los dos momentos coinciden o no.

De esta manera, con una lógica temporal ockhamista enriquecida con nominales, podemos avanzar en el análisis del momento *t* en el que un punto de vista se atribuye a un sujeto.[4]

REFERENCIAS

Areces, C., Blackburn, P. y Marx, M. 2001, 'Hybrid Logics: Characterization, Interpolation and Complexity, *The Journal of Symbolic Logic*, 66(3): 977-1009.

Blackburn, P. 1994, 'Tense, Temporal Reference and Tense Logic', *Journal of Semantics*, 11:83-101.

Blackburn, P. 2006, 'Arthur Prior and Hybrid Logic', *Synthese*, 150:329-372.

Blackburn, P. y Goranko, V. 2001, 'Hybrid Ockhamist temporal logic', en Bettini, C. y Montanari, A. (eds.) *Proceedings of the Eight International Symposium on Temporal Representation and Reasoning*. TIME-01, Cividale del Friuli, Italy.

Liz, M. (ed) 2012, *Puntos de vista. Una investigación filosófica*, Barcelona, Laertes (en prensa).

Ponte, M. y Vázquez, M. 2008, 'An approach to the notion of point of view through hybrid temporal logic', *ESPP 2008*, Utrech.

Ponte, M. y Vázquez, M. 2012, 'Tense and Temporal Reference. Hybrid Temporal Logic', *Logique et Analyse*.

Prior, A. 1957, *Time and Modality*, Oxford, Clarendon Press.

Prior, A. 1967, *Past, Present and Future*, Oxford, Clarendon Press.

Prior, A. 1957, *Papers on Time and Tense*, Oxford, Clarendon Press.

Vázquez, M. y Liz, M. 2011, 'Models as Points of View. The Case of System Dynamics', *Foundations of Science* 16(4):383-391, Springer Verlag.

[4] Este trabajo ha sido financiado con los proyectos FFI2008-01205. Puntos de Vista. Una Investigación Filosófica y FFI2011-24549. Puntos de vista y estructuras temporales.

SEMANTICS AND EPISTEMOLOGY OF ARTIFACTUAL KINDS

Coordinators:
JESÚS VEGA-ENCABO
Universidad Autónoma de Madrid

VÍCTOR M. VERDEJO
Universidad de Santiago de Compostela

Introduction

For a long time, artifacts have been notably neglected in the philosophical literature. However, they raise substantive and distinctive problems and puzzles. This symposium discusses a number of topics regarding the epistemology of artifacts and the semantics of our terms of artifactual kinds. Its main objective is to extend recent discussions on the metaphysics of artifacts *qua* artifacts and the nature of artifactual kinds to epistemological and semantic issues.

Not only do our sortal terms for artifactual kinds reflect our practices to classify and categorize objects in the environment, but also they seem to help us to pick up particular artifacts by their belonging to genuine real kinds. However, do sortal terms for artifactual kinds work like our terms for natural kinds? And how do we ground knowledge we acquire from artifactual kinds? Is this knowledge an outcome of scientific discovery?

Artifacts are a kind of objects made by us. One way of giving some metaphysical import to this fact is by accepting an intentional theory about the nature of artifacts. The main idea is that the identity of the kind an artifact belongs to is fixed by the intentions of the makers. In a series of recent papers, Amie Thomasson has argued that this fact entails certain theses about the way in which we are semantically and epistemically related to artifact kinds which strongly contrasts with common assumptions about natural kinds (Thomasson 2003, 2007). Firstly, we apparently exhibit a sort of privilege regarding our knowledge of artifacts, one that is understood as involving immunity to massive error and the impossibility of being deeply ignorant about them. These two features could, secondly, undermine some classical arguments for an externalist semantics of artifactual kind terms: the so-called arguments from error and ignorance. The contrast with common assumptions in the epistemology and semantics of natural kind terms is explicit and evident.

The topics discussed in the four presentations of this symposium bring about the difficulties to make sense of some theses around the authorship of artifacts, the epistemic privilege (at least, of the makers) and an externalist framework in semantics. The aim is to provide central insights into the semantics and epistemology of artifactual reality.
These are the main issues to be discussed:

1. Is there a sense in which a subject can be said to know what an artifact is in virtue of being its maker? Does it make sense to talk of a practical-intentional knowledge of the artifacts as the products of our making? Are there different kinds of knowledge we can have of artifacts (e.g., the ones based on observation, practical interaction or authorship)? If so, how are they interrelated?
2. How is the thesis that an artifact is the result of the intention of an author/maker related to the apparent epistemic privilege the maker exhibits regarding the products of its making? Should we assume that both these theses are necessarily connected?
3. Should the contrast with natural kinds be yet maintained once we abandon certain intuitions about our knowledge of artifacts?
4. How to defend externalist views on the semantics of artifactual kind terms against the arguments of the descriptivist? What argument does the externalist need in order to rebut some semantic intuitions that are apparently grounded on the metaphysical mind-dependence of artifacts?
5. Do people really use cues about the intentions of the makers to ground their categorization judgment about artifacts? What sort of judgments is supposed to underlie our grasp and mastery of artifactual kind concepts? How are our categories influenced by cultural variables in the case of artifacts?

ARTIFACT CONCEPTS AS CULTURAL COORDINATES

MARÍA MUÑOZ SERRANO
Universidad Autónoma de Madrid

ABSTRACT: It is commonly accepted that our categories to classify our world of artifacts cannot be grounded in the identification of the objects according to a material or formal essence. In the light of this fact, two kinds of theories have been proposed to understand artifact categorization: those that identify artifacts on the basis of their functions and those that do so on the basis of the intentions of their maker. In this paper, I will argue that neither of them is sufficient to elucidate the way in which human subjects are able to categorize and conceptualize artifacts. My proposal is that classifying an artifact depends on the coordinates that it occupies in a cultural space. That leads us to a dynamic and relational approach to the practices of categorization of artifacts. Finally, I would like to remark that this idea allows us to set a parallelism between concepts and affordances of artifacts. Firstly, because both notions share a dynamical and relational character. Secondly, because we can understand the concepts of artifact as linguistic correlates of such affordances that help us to fix and constrain the meaning of an artifact.
KEYWORDS: artifact, cultural coordinate, category, affordance, human practices

I

It is commonly accepted that there is a difference between our ways of categorizing natural elements and artifacts. The first sort of categorization seems to be established on the basis of membership to a natural kind. In the case of natural things, categorization and kind membership are determined by the possession of a range of mind-independent properties (physical, chemical, structural properties, etc.) of the object. However, it seems intuitive that this is not the case with artifactual categorization; material properties of a screwdriver, such as its chemical composition or physical structure, may be relevant, but they are not crucial to identify a screwdriver as such. Artifacts are created by agents, they are products of their intentional activity; so we can assume that their categorization has to be somehow related to those agents' minds and action routines and we cannot just pay attention to some kind of ontological features as the basis for such a categorization.

A possible solution for the problem of artifactual categories is to consider the function as the central element for ascribing a token to a specific group. For instance, every object whose function is bringing small quantities of liquid to the mouth could be grouped in the category “spoon” or named “spoon”. Although this approach seems to be coherent with our intuitions, there are very well-known objections to it. (1) Sometimes artifacts do not have a clear function, this is the case of artworks or ornaments, consequently in these cases considering their function would not be enough to assert that the artifact in question is a member of a certain category (Thomasson, 2007, 57). (2) Some objects have the same function, but we categorize them as belonging to different kinds; we could say that a vacuum cleaner and a scrubbing brush share at some level of description the same

function, but rarely put them into the same category. Of course, it is possible to be more specific about the function of those objects and redefine it conveniently. For instance, the function of the vacuum cleaner is to suck up dust and to store it, a scrubbing brush is for dragging it. That poses a problem for a robust ontological functional theory of artifacts that would provide different groupings depending on the level of description we use to identify properties, given that nothing can secure that these different levels provide a clear hierarchical structure. But at the same time this weakened notion of function will give to us a very confusing and changing system of categories (3) Finally, there are broken or defective artifacts lacking in function that are considered (broken or defective) members of a category; this seems to suggest that what is really important is not the actual function, but the intended one, the function for which the artifact was made.

Now, if we admit that categorization is done on the basis of the intended function, then the focus from which we view the categorization process changes from the artifact itself to the maker, the person that intended that the object had a specific function. This leads us to take into consideration the possibility that the ascription of an object to an artifactual kind could be dependent of the maker's intentions (the intention of copying an object of that kind or, in the case of the maker of a prototype, the intention of creating an artifact with a specific function). Amie Thomasson proposes (Thomasson, 2003, 2007) that, in fact, functions are not crucial to determine kind membership in the case of artifacts, but the maker's intentions (independently of whether those intentions fix the intended function of the object). Hence, there is something like a baptismal act when an agent creates the first token of a category, her intentions determine what means to be an object of that kind, and doing so the maker is immune to some kinds of categorization error. However, again, some objections threaten the plausibility of this proposal. (1) At first sight, this is an individualistic approach that assumes the existence of a singular original maker with a singular and concrete intention. But understood in these terms, it fails to explain cases of cooperation or collective creations, where maybe the object produced is not consequence of the intention of a singular agent, but result of the interaction between different agents, with different intentions, none of them squaring with the final product. Even if we get a satisfying explanation of how agents can form those collective intentions, we could argue that those intentions (collective ones and personal ones) are shaped by other social elements such as social rules, conventions, institutions, other artifacts, etc. We have then to take into account those social elements in order to explain maker's intentions and, therefore, membership to an artifactual kind (Preston, 2006) (2) We can also wonder if the idea of an agent creating a completely new artifact has any grip in reality, given that the most part of the objects around us are the result of a historical process of technological change. It is clear that most of artifacts are descendants of previous objects and that, during this historical process,

their functions and properties are changing (Basalla, 1988). Even if we could find a case of absolute originality, we can wonder if it is possible to posses the intention of creating that original object, that is, if it is possible to posses the concept of an object that does not yet exist (Vega & Lawler, in press).

I think that neither appealing to function nor appealing to maker's intention is enough to ascribe a particular object to an artifactual kind. I will assume some premises as the basis of my paper: First, there are different important elements for determining if an artifact belongs to a specific category, such as physical properties, historical context, function (of course), etc. Second, we cannot reduce our practices of categorization or related to categorization (naming, grouping, etc.) to just one, neither expect that those practices could be explained in the same way (Malt & Sloman, 2007). Finally, we need to assume that artifactual kinds or categories are flexible and that an object can belong to different categories at different moments or as result of different categorization practices (Barsalou, 1987).

II

Imagine we are in one of those huge clothing stores, part of a multinational chain, where all the products are arranged in different aisles (women, men, kids, sport clothes, underwear, etc.). We find a pair of black cotton trousers in an unexpected context, in the shoes aisle for instance. What kind of trousers are they? By their properties (shape, material), they could be pyjamas or sport trousers, and probably we would decide their category paying attention to the aisle where they are: if the trousers are surrounded by nightclothes, robes and underwear we will easily categorize them as pyjamas; but if we found them close to the yoga mats and sneakers, we will be more prompted to consider the trousers as sportswear; if the trousers are among rain boots, sandals and pumps, it is clearly ambiguous.

My proposal is that categorization in the artifactual realm works analogously to that store arrangement. Artifacts rarely are isolated from other objects and spaces; the artifacts we interact with do not use to be found in neutral, or even unexpected, places. I am looking at my desk, and there are a lot of different objects: there is a laptop, some books, some markers, a pencil holder, some folios, etc. But the point is not how many artifacts I can find on my desk, but how they are related and how they do depend -even for their existence- on each other. The markers and the pencil holder have indeed a relation (not surprisingly the pencil holder is containing the markers). If we do not take into account that relation we cannot say what a pencil holder is, because it does not make sense isolated of artifacts such as markers, or pencils, or crayons (in fact the pencil holder is a

former jam jar). The ascription of an object to the category "pencil holder" is connected to the position of that object with respect to some other artifacts that we categorize as pencils (or markers, or pens). But at the same time it exists a relation between markers and folios, we cannot consider something as a marker if it is completely separated from any surface susceptible to be painted (paper, whiteboard, etc.), and probably we could consider paper sheets something really different if they were not related to artifacts for writing (markers, but also pencils, typewriters, etc.). In conclusion, it seems that artifacts have meaning for us not only in virtue of some particular properties (physical or even functional) of the object, but also based on the relations they hold with other objects in a rich artifactual environment. Those relations may be hierarchical, like the relation between pencils and pencil holders, but they are not necessarily hierarchical: for instance, paper sheets and markers seem to be mutually dependent. Therefore, we can say that relations of different kinds conform a complex network of artifacts, where meaning, function or history of an object are supported by others artifacts around it.[1]

That network partially shapes the *space* of my office. An office is a really concrete context with very specific objects: computers, desks, chairs, books, a printer, etc. There hardly are unexpected artifacts in this context, as well as artifacts do not use to appear in unexpected contexts. This context invites somehow to certain practices and makes some actions more difficult. It is a perfect place for reading, writing, stapling documents and checking facebook, but it is not a suitable place for picnics (despite some officemates opinion) or table tennis. Artifact networks influence the context where we act and the practices that we consider available, but at the same time, those practices condition the introduction of new artifacts in that context and the structure of the network of artifacts.

I will call those positions that artifacts occupy within artifact networks and contexts, cultural coordinates. These cultural coordinates specify a place in the cultural space in relation with other objects, practices and agents. It is important to highlight, on the one hand, that an artifact cannot be reduced to its cultural coordinates, because we cannot reduce all its properties, for example material properties, to that position; but that position determines the meaning, the value and the category of an artifact within a community. On the other hand, neither artifactual categories can be identified with cultural coordinates; coordinates are specific points in a cultural space, categories cover wide parts of that space, different coordinates fall under the same category and their limits are fuzzy; so there is a covariance between cultural coordinates and artifactual categories, but it does not mean that they are the same.

[1] We can find some similarities between this idea and Heidegger's analysis of the surrounding world in "Being and time" (Heidegger, 1967).

My thesis is that category membership, in the case of artifacts, is fixed by the cultural coordinates that an object occupies, that is, its position relative to those relevant elements that have been mentioned. When a particular artifact changes its coordinates and is put in relation with new elements or situated in a new context, then the concept under which that token falls could change. This is the case of the jam jar that I brought to the office and it became a pencil holder. So, different changes can imply new cultural coordinates, changes in context, in relations with other object, or the introduction of new practices or new functions. But none of these changes are necessary for situating an artifact in new coordinates and, hence, under a new category; that is, we can consider that an artifact is keeping its function (at some level of description), but its context has changed. It could also happen that an artifact keeps its context, but changes its function. In both cases, they are occupying new coordinates in the cultural space.

On the other hand, changing the position of an artifact in the cultural space does not always lead us to fix a new concept for that object. Sometimes, for instance in the case of the black cotton trousers, the coordinates do not determine the category, but introduce some grade of ambiguity when we are involved in categorization practices, probably because the artifact is situated in a context where the relations with other objects and actions are weak or meaningless. There are some cases of re-situation that do not imply a definitive and stable change of category, but a temporal or partial re-categorization; if I do not have a screwdriver, but I want to build my new Ikea shelves, I can use a knife to do it, and maybe I can store the knife in the toolbox until I find a real screwdriver. Is that knife a screwdriver now? My intuition is that it partially is. If someone ask me for a screwdriver, probably I will say “I do not have one, but you can use this knife”, and if I need to repair a piece of furniture I will not hesitate to use the knife; but, probably, I will consider it a knife and, if I find a real screwdriver, I will return the knife to the kitchen, to its habitual cultural coordinates.

However, I would like to defend that this flexibility does not entail a subjectivist approach, where categorization only depends on the agent's judgments. The position that an artifact occupies within these cultural coordinates is socially and historically determined and shared by a whole community (probably there will be differences of degree: some cases can be more easily accepted than others). There are also cultural constraints that prevent to change that position: a specific object cannot be re-situated in any position; and some of these changes will be more plausible than others: it is easier that a knife become a screwdriver than a replacement for my dental floss. In addition we have to take into account that not every change in the cultural space has the same status, some changes will be accepted only for a few member of the community or just as temporal situations,

but stable changes on the cultural spaces and permanent ascription to a new category cannot be a result of an individual judgment, but an event within a community.

III

In this last section I would like to explore the relation between artifact affordances and ascription to an artifactual category. First, I will briefly clarify the concept of affordance, what will allow us to set a comparison between both notions.

The notion of affordance belongs primitively to ecological psychology (Gibson, 1979) and refers to possibilities of action that can be directly perceived by an agent in her environment. For instance, a human being can perceive a stone as graspable or *throwable*, whereas for a spider the same stone can be perceived as climbable or *dwellable*. These affordances are not exactly properties of the object in the environment, nor properties of the agent, they are relational properties that emerge from the interaction between agent and environment.

However, this notion does not seem enough to explain our practical access to artifacts. Indeed we can assert that artifacts are perceived as possibilities for action, but we need to explain why some affordances are more salient than others, why artifacts have some possibilities for action strongly associated. I can use one of my markers to stir the coffee or as bookmark, but those uses seem to be less relevant and less often perceived than the possibility of using it for writing or painting. These salient affordances of artifacts have been called canonical (Costall, 1995) or intentional (Tomasello, 1999) affordances[2].

Canonical affordances specify not only a possibility of action, but a privileged use, that has been culturally selected and determined, by means of practices, habits, relations with other environmental elements, etc. Therefore, canonical or intentional affordances emerge not only from the interaction between an agent and the environment, but from the interaction of agent, environment and cultural context.

So, what is the relation between affordances and concepts of artifact? I hold that, first, there is a parallelism between possessing a (canonical/intentional) affordance and falling under a concept of artifact; and second, that we can understand artifact concepts as correlations of canonical or intentional affordances. Both notions (artifactual category and canonical or intentional affordance) share some principles as consequence of their dependency of the cultural space (Neisser, 1987): (1)

[2] Of course, there are some differences between both notions (canonical and intentional affordances), but the distinction is not relevant for my aims in this paper.

They are both relational notions and at the same time they can be both considered as the result of complex interactions; (2) in addition, as much as they are relational, they are dynamical and flexible, because they are subject to changes depending on those interactions; (3) finally, both are dependent of cultural processes and contexts that are part of these interactions: practices, spaces, cultural meanings, etc.

We can consider artifact concepts as correlates of the canonical/intentional affordances. If an object is a member of an artifactual category, it will mean that there is a specific interaction with the object, that is, that the artifact has a function, but also a role within a cultural space that connect the artifact with other elements of that space in special ways: the artifact has a privileged use that cannot be identified with its function[3]. A final reflection: maybe our categorization practices are a strategy to fix the meaning an artifact acquires within a community and to find it a place within a rich environment of other objects, practices and actions. The canonical/intentional affordances allow us to situate artifacts in our cultural space, but probably relying just on the affordances of the object would give to us a too weak and flexible placement. So categorization will provide us with a way to explicitly fix that placement (or positioning), improve the relevance and salience of the privileged use and constrain the possible interactions with an object.

Acknowledgments

I thank Jesús Vega and Álex Díaz for their comments and suggestions. This work is part of the research project "Epistemology of artifacts: Affordances, practical knowledge and epistemic artifacts" (FFI2009-12054) and have been funded by the Ministerio de Economía y Competitividad (Spain).

[3] The distinction between function and privileged use is not completely clear, my intuition is that we cannot assume that they are the same, because "function" is a teleological notion, but a privileged use is not necessarily purposeful.

THE INTERNALIST/EXTERNALIST DEBATE ABOUT ARTIFACTUAL KIND TERMS

Laura Ortega
Universidad de Barcelona

ABSTRACT: The debate between externalism and internalism concerning artifactual kind terms begins with the Putnamian suggestion that artifactual kind words, as 'pencil', have an externalist semantics in the very same way that natural kind terms, as 'water'. The argument offered by Putnam is rejected for both critics and defenders of the Putnamian conclusion because Putnam does not characterize properly the relevant nature of artifacts. All parts in the debate agree that the relevant nature of artifacts is its function, not its origin, and that externalism needs arguments of ignorance or error concerning the function of artifacts. Nelson and Kornblith present such arguments. However, I think an internalist can reply those arguments because they are not *universal* arguments of ignorance and error, leaving untouched a modified version of descriptivism. Is, then, such a universal argument available for externalism? I think it is. And although descriptivist replies to that argument cannot be excluded, I think is a type of argument that puts more pressure on the side of internalism.
KEY WORDS: Externalism, internalism, artifactual kind terms, argument of ignorance and error

Putnam proposes that the externalist semantics that he defends for natural kind terms, such as 'water' or 'gold', can be extended to artifactual kind terms, such as 'table' or 'pencil'. That is, the reference (or domain of application) of such terms is also determined, as in the case of natural kind terms, by something *external*: the possibly unknown relevant nature of the members of the kind. He argues for this externalist position for natural kind words appealing to the well known Twin Earth thought experiment. In the case of artifactual kind terms, he offers the argument of pencils-organisms: imagine that we discover that all pencils in Earth are really organisms and not artifacts, as we believed. We cut them, open them and see that they have organs, tissues, nerves, etc. (Putnam 1975, 131-93). In such a possible situation, Putnam says, it would turn out that pencils are not artifacts made for writing, as we supposed, but organisms that we use for writing. Thus, we could be mistaken about what the relevant nature of pencils is, and that possibility shows that the extension of the term 'pencil' cannot be determined by the description we associate with the term, 'artifact made for writing'.

This Putnamian argument, however, has not been as convincing as the Twin Earth argument for 'water'. The problem with this argument is that Putnam presupposes that the relevant *nature* of pencils we could be mistaken about is *compositional* or *original*, instead of functional. In this line, Schwartz concedes that pencils could turn out to be organisms but replies to Putnam that the description we associate with 'pencil' is not '*artifact* made for writing', but some neutral description about the artifactuality of the things that actually are artifacts, some description like

'*thing* used for writing (*with such and such form*)' (Schwartz 1978, 566-74). That is, the relevant nature of artifacts, whose descriptive specification we attach to the corresponding term, is not compositional or about its origin (contrary to the case of natural objects), but mainly functional. According to Schwartz, then, the Putnamian argument does not work, and nothing rules out that the reference of 'pencil' is determined by the description we know about pencils: 'objects (*with such and such form*) used for writing'. So, an externalist for artifactual kind words must appeal to some argument showing that we could be ignorant or mistaken about the function of the members of an artifactual kind in spite of referring with the corresponding artifactual kind term. That is, externalism needs arguments of ignorance and error concerning the function of artifacts. This starting point has been accepted both for critics and defenders of Putnam's view with respect to artifactual kind terms.

After the Putnamian suggestion, some thought experiments, showing that we could be ignorant or mistaken about the function of an artifact, have been presented in favour of an externalist semantics for terms for artifacts. Thus, Kornblith proposes the following thought experiment: let us suppose that Martians arrive to Earth and want to investigate human artifacts scattered all around (knowing solely that those objects have been made by human beings). A Martian anthropologist, ignoring the function of doorstops, finds a doorstop, points to it, and say 'Let's call the kind of which this is a member 'glug''. In spite of his ignorance about the function of doorstops, his reference to doorstops using 'glug' seems successful (Kornblith 1980, 109-14). And, according to Kornblith, the Martian ability when using 'glug' needs not to be parasitic upon the existence of human experts on doorstops. We can suppose that 'glug' is a term only introduced in the Martian language and, therefore, no human being associates a description with the term. Or even we may suppose that Martians visit the Earth after the disappearance of mankind and, thus, there would be no experts in doorstops. In such a situation, any description that the Martian anthropologist associates with 'glug' is insufficient to pick up the domain of application of the term. Thus, against internalism, Kornblith concludes that it is possible to introduce and use a term for an artifact without knowing any specification that individualizes the artifactual kind in question.

The second externalist argument is proposed by Nelson. It is an argument based on error that invites us to think that we discover that pencils are not really instruments for writing, as we had taken them to be, but alien implements for spying and control. Suppose that the objects we call 'pencils' are really instruments that evil aliens have scattered about the Earth in order to manipulate human activities. Pencils are really mechanisms by which aliens psychologically control us and trigger our wars. In order to disguise their intentions, Martians make us believe that pencils serve to write or draw, but they never had such a function. Pencils just produce on humans the illusion that

they serve to write or draw (Nelson 1982, 359-65). If such a situation is epistemically possible, then it shows that humans could have been applying the term 'pencil' without a correct description of the function of pencils. So, concludes Nelson, the extension of our term 'pencil' could not be determined by any cognitively accessible description that humans have in mind. Reference would be determined by the functional nature of pencils, whatever it is. Nelson admits that the plausibility of his example depends on the plausibility we are disposed to concede to such thought experiments. But appealing to the collective illusion is a mechanism always available:

> "After all, if Putnam can hold that pencils are organisms, despite the fact that large numbers of people would give their occupations as pencil-makers, we can tell our story as well." (Nelson 1982, 362)

In sum, if the thought experiment is coherent, the descriptivist could not identify the alien instruments as pencils, since they do not fit the description we associate with the term. In contrast, for an externalist, such implements are pencils; we simply were wrong about its function. As in the case of natural kinds, what we believe about the relevant properties of artifacts (the function and, perhaps, some other properties) is epistemically corrigible.

These two externalist arguments intend to show that reference of artifactual kind words is possible even though nobody has a singularizing description associated with the term, due to ignorance or error concerning the function of the artifacts in question. In the first case above, the Martian anthropologist neither knows any singularizing description associated with 'glug' nor he can defer to any expert who knows such a description, but it seems that he successfully refers when using the term. In such a situation, a descriptivist cannot appeal to deference to human (or expert human) uses of 'doorstop' because deference demands a historical causal connection between uses or speakers. And, in the argument, it is obvious that Martians and Humans belong to different linguistic communities; their respective uses of 'glug' and 'doorstop' have no connection between them. So, in the absence of a causal chain, reference cannot be transmitted from Humans to Martians. Kornblith's argument show that an artifactual kind term can be introduced and used without any expert associating with the term a singularizing description of the members of the artifactual kind. Rather, the situation of Martians with respect to 'glug' is analogous to *our* situation when we introduce a natural kind term. In the case of Nelson's argument, it seems that human reference to pencils is successful in spite of the human mistake about the function of those artifacts. Again, Humans in such a situation could not appeal to alien experts on pencils because our linguistic practices would be disconnected from theirs. No historical chain would connect human and Martian uses of 'pencil'. In sum, these two arguments intend to show that it is possible to introduce and/or use a term for artifact with speakers' ignorance and error about the relevant

properties of the artifactual kind in question. The externalist general conclusion is that Putnam was on the right track and the descriptivist must concede that there is a greater similarity than expected between natural kind terms and artifactual kind terms.

Now I consider two descriptivist replies to those externalist arguments. One line of answer consists in saying that, in the mentioned cases, reference is unsuccessful. Thus, in the case of 'glug', Martians do not refer to doorstops when they use the term 'glug' because of the lack of an associated description with the term. The term does not pick up anything. In the case of spying pencils, humans do not refer to pencils, simply because there are no pencils (that is, instruments for writing) at all. So, after all, a descriptivist could say, Nelson's thought experiment is not plausible: we could not discover that pencils are not entities for writing or drawing simply because, in the described situation, there would not be pencils. This is the line of response that Schwartz and Thomasson endorse. Another possible answer to Nelson and Kornblith's arguments would consist in saying that those cases are not the common cases of use of artifactual kind terms. Perhaps, in those peripheral cases, reference can be determined in an externalist way, but this does not happen in central cases. This point of view seems to be suggested by Thomasson in some occasion. Typical cases of introduction and use of artifactual kind terms are cases in which the community that introduces and uses an artifactual kind term is the same community that creates the corresponding artifacts. In these standard situations, a descriptivist could say, the reference of the terms of artifacts is determined by the corresponding associated descriptions known, at least, by the makers or experts on those artifacts. Thus, although the Martian term 'glug' and the human term 'pencil' may have an externalist semantics in Kornblith and Nelson's argument, respectively, this could not be the case for 'doorstop' in the human community or for the alien term for their spying instruments.

I think, however, that these possible internalist responses are unsatisfactory. It seems that, in both mentioned situations, reference to pencils by humans and reference to doorstops by Aliens, respectively, are successful. Especially in Nelson's thought experiment, it is hard to admit that all past human uses of 'pencil' had no referent at all. But, of course, if the issue is one about externalist or internalist intuitions concerning the success in referring, then the issue remains undecided. Schwartz thinks that, in Nelson's example, we would describe the discovery of the function of *pencils* as a situation with no *real* pencils. But, I think this is not so clear. It also seems plausible to describe the situation as one in which we just were wrong about the function of pencils. Appealing to the intuitions about what we would say is not crucial, since intuitions can be divided and different intuitions may seem plausible.

On the other hand, if descriptivism admits that reference is successful in these (bizarre) examples,

then descriptivism cannot maintain that artifactual kind terms can *never* be used successfully without a cognitively accessible associated description (or a deference to such a description) picking up all and only the members of the artifactual kind. I take that the descriptivist thesis for artifactual kind terms is something along these lines: in order that a speaker can successfully refer when using an artifactual kind term, it is necessary that she has cognitively access to the description that picks up the artifactual kind or that she defers to someone who knows that description. Then, the externalist arguments would show that descriptivism fails because reference to artifacts is not *necessarily* determined in *every* case of use of an artifactual kind term by an associated specification known by someone. At this point, it is clear that the previous externalist arguments exploit the possibility that a term for a certain kind of artifact can be introduced and used by a linguistic community which has not created the artifactual kind in question. Similar arguments to that by Kornblith can be built appealing to the discovery of archaeologists or anthropologists of unknown artifacts from other cultures or epochs. And similar arguments to that by Nelson can be built appealing to the fact that the *real* function of a certain type of artifacts could be unknown for some community (as a result of an illusion created by different intelligent beings). But these possibilities, made explicit by the externalist arguments, had not been taken into account by internalism. So, if an internalist is going to defend that descriptivism explains well how terms for artifacts work in common circumstances although in marginal situations externalism can be applied, I think the best she can do is to modify its thesis, relativizing it to linguistic communities. She could say something along these lines: for every artifactual term, there is a singularizing description of the kind associated with that term in *some* linguistic community -presumably the community that introduces the artifact- and, at least, in that community, the reference of the uses of the term in question is determined by that description known by someone. This internalist thesis, relative to linguistic communities, could resist the externalist arguments. In fact, according to Thomasson, the externalist arguments only show that *some* linguistic community might be mistaken or ignorant about the relevant nature of some kind of artifacts (Thomasson 2003, 580-609). The point is that the ignorant or mistaken community is not (and, presumably, cannot be) the community that introduces or creates the artifact in question.

Thus, externalism would need a *universal* argument of ignorance or error affecting *every* community, including the community that makes the artifact. Is available such an argument? The difficulty lies in imagining that some community -humans, say- *makes* or uses some artifact, with an attributed function in that community, which really has another function unknown for that community. The problem is that having *another* function demands another intelligent being different from us, and, then, we would have the same type of argument as the previous ones, in

which there is some community – the producers of the artifact- that knows the distinctive properties of these types of artifacts. So, in order to get a universal argument of ignorance or error, an externalist cannot appeal to another function *for someone else* and unknown for us. But I think an argument of universal error is possible if we imagine that we could be mistaken about the attributed function to a particular kind of artifact (without a new function to discover). Consider, for instance, that we discover that soap powder we use to wash dirty clothes really does not clean at all: spots really continue there after *washing* the clothes. Due to some properties of the ingredients we use to make soap, affecting our perception system, we have been all the time under the optical illusion that spots disappear under the soap. In such a situation, our description associated with 'soap', 'entity that serves for cleaning the spots', would not be a true description of the artifacts we called 'soap'. All we would have been wrong, including makers and experts on soaps, when thinking that those samples we call 'soap' have the function of cleaning. But, even so, it seems that our uses of 'soap' have been successful in referring to those samples we called 'soap' and that really make us suffer from certain optical illusions.

If the imagined situation is epistemically possible and reference is successful, the requirement of a universal argument of ignorance or error can be satisfied. Of course, a descriptivist may reply, for example, that in such a situation, if possible, Humans have never referred to anything with their use of 'soap' because soap never really existed. But I think that this objection is not satisfactory; human past uses of 'soap' are inserted in human linguistic practices and some explanation about the reference and meaning of such a term must be given. Perhaps internalism may have different objections to that argument of universal error, but I think this kind of argument poses a more pressure situation for descriptivism.

Acknowledgments

This paper has received the financial support from the project "Reference, self-reference and empirical data" funded by the Spanish Ministerio de Economía y Competitividad (FFI2011-25626), and from the grant FPU (AP2008-00909) of the Spanish Ministerio de Educación, Cultura y Deporte.

TWO KINDS OF KNOWLEDGE OF ARTIFACTS

JESÚS VEGA ENCABO
Universidad Autónoma de Madrid

ABSTRACT: In this paper I apply a distinction proposed by Stuart Hampshire about two kinds of knowledge of the future to our knowledge of artifacts. I respectively call them: practical knowledge and inductive knowledge of artifacts. I will argue that maker's knowledge of artifacts could be understood in principle as a kind of practical knowledge that could also make sense of some of the features that we associate with the sort of epistemic privilege we have regarding artifacts and artifactual kinds. Nevertheless, I will criticize some ways of understanding this suggestion and I will offer an epistemological model of our knowledge of artifacts that claims the following: the conception that guides the production of the artifact is sufficient to produce the artifact through a reliable mechanism that is under the control of the subject. And this is only possible if the subject is able to identify a set of success conditions learned within a dynamics of exchanges with an environment rich in functions and purposes.
KEYWORDS: practical knowledge, maker's knowledge, artifacts, epistemic privilege.

1. Two kinds of knowledge: makers and non-makers of artifacts

Consider the following text by Maimon:

> "THERE is a great difference between the knowledge which the producer of a thing possesses concerning it, and the knowledge which other persons possess concerning the same thing. Suppose a thing is produced in accordance with the knowledge of the producer, the producer was then guided by his knowledge in the act of producing the thing. Other people, however, who examine this work and acquire a knowledge of the whole of it, depend for that knowledge on the work itself" (Maimon, *Guide for the perplexed*, part III, chap. 21)

Maimon, in his *Guide for the perplexed*, argued for a distinction between two kinds of knowledge regarding the products of the activity of making. The distinction is introduced in order to justify an analogy between the knowledge God has of the creatures and our knowledge of nature. The very same things can be known in two different ways, one of them grounded on the very fact of these things being products of our making (on God's making). Let us call this kind of knowledge *maker's knowledge*; the other, grounded on observation and inference, *non-maker's knowledge*. Some features allow us to distinguish between both kinds of knowledge of what we can call broadly speaking *artifacts*. Let us consider first *maker's* knowledge:

1. Maker's knowledge is the kind of knowledge that guides the producer in the process of bringing about the artifact; it is a kind of knowledge that regulates and controls the

intentional production of the object. Besides the artifact is made in accordance with the knowledge in a normative, rule-like way (not merely causal).

2. This kind of knowledge does not depend on the object produced; it is in a sense prior to it. Nevertheless, it is yet knowledge about the empirical working of the artifact itself.

3. Though in a sense "empirical" knowledge of the artifact, it is not grounded through the observation of how the artifact functions (plus some inferences). The "truth" grasped by the maker depends on the fact that "the movements be produced in accordance to his knowledge", as Maimon would say.

4. Maimon also suggests that makers have a complete understanding of the thing produced.

5. Maker's knowledge consists in a capacity to anticipate the future and the non-actual working of the artifact.

Contrast with non-maker's knowledge:

1'. A non-maker is not guided by previous knowledge of A; he/she obtains new information about the artifact through the continuous interaction with it.

2'. It entirely depends on the object; obviously, it represents knowledge about the empirical working of the artifact.

3'. It is grounded on the interaction with the object through an observational process. The "truth" grasped by the non-maker is yet about the artifact itself.

4'. Non-makers can also acquire through observation a *complete* understanding of the working of the artifact.

5'. But there are some limits to the grasp of what an artifact is: i) our observations are finite, in a way that it would make impossible to acquire complete knowledge when infinite movements are involved; ii) knowledge acquired is always dependent on actual and occurrent conditions.

In brief, the main characteristic of maker's knowledge is that what is known is known, in a sense, by being the outcome of the knowledge of it (what at first sight seems paradoxical), whereas in non-maker's knowledge what is known is known by a process of "extensive interaction" with what it is known. The very same facts (if we are talking here about factual knowledge) that are known by the maker could be known in this way by non-makers. It is as if in the first case, when we talk about maker's knowledge, it is not the artifact itself the standard against which we measure the correctness, but the knowledge itself that serves as a measure for what the artifact is, whereas our

ordinary empirical knowledge of the artifact, as an object among other objects in the world, measures its correctness or incorrectness against the artifact itself.

2. Two kinds of knowledge: practical and inductive knowledge

Stuart Hampshire, in his book *Freedom of the individual*, proposes the thesis that there are also two different kinds of knowledge of the future: the *practical knowledge* of what will happen because we have the intention to do it and the *inductive knowledge* of what will happen grounded in an inductive reasoning that make some events expectable given what has happened till now (Hampshire, 1975). What *I* will do (my actions) could be known in these two different ways. One needs to resist the temptation of explaining the difference in terms of their respective possible objects, as if what is known practically is not known inductively and viceversa. I can know my actions (I would say, present and future) in both ways: one of them grounded in my intending to do it and the other grounded in an observational process. Stuart Hampshire's idea –and in this he concurs with Elisabeth Anscombe- is that *the same thing*, e.g. an action, could be known in two different ways, even if we cannot acquire both kinds of knowledge at the same time, because each depends on adopting a certain attitude, intentional or agential in the first case, observational or theoretical in the second.

Though my introduction of the terminology of maker's and non-maker's knowledge could be thought in terms of different individuals having different kinds of knowledge of an artifact, we could say that in some respects the same subject can acquire both kinds of knowledge. In the same way which the difference between both kinds of knowledge of our actions cannot be resumed in a distinction of their respective objects, so the difference between what I have called maker's knowledge and non-maker's knowledge cannot be understood as a difference in the knowing subjects. The thesis is that they are, in principle, two different *kinds* of knowledge that demand separate epistemological conditions. I claim that the distinction introduced by Maimon can be understood in the terms of different attitudes towards what it is known, the artifact or the product of the making: the first attitude concerns the agential involvement in the making of the artifact; the second attitude has to do with the acquisition of information about the object through empirical ordinary means, that involve observation, inference, and also testimony. Maker's knowledge of artifacts is a sort of practical (intentional) knowledge. Non-maker's knowledge of artifacts is a sort of inductive knowledge.

This parallel suggests that artifacts belong with intentional actions to this part of reality that can be known in two different ways. One interesting suggestion can be that part of reality can be empirically known only insofar as it is viewed under the kind of intelligibility that is linked to our

agential involvement in the world. Because we can have some sort of practical knowledge of artifacts, we can also acquire some empirical inductive knowledge of them. This is not surprising given our characterization of maker's knowledge, because in a sense what an artifact is depends on the fact of being produced through the guidance of practical knowledge of it. I have to put aside here the metaphysical question behind this idea. But what it is really crucial is the task of providing an epistemology of practical knowledge that could be applied coherently to the knowledge not only of the intentional actions (as in Anscombe and also in Stuart Hampshire), but also of the products of these intentional actions, i.e. artifacts.

Maker's knowledge shares then the very same epistemological difficulties as practical knowledge. Firstly, it exhibits features that are not in principle characteristic of our ordinary empirical knowledge. Secondly, it is a kind of knowledge that is not governed by the object -by what is known- and for that reason it does not apparently answer to the facts. Thirdly, in a more positive vein, this kind of knowledge regulates and controls the intentional production of the object, what it means to revert what it could be called the *control direction* characteristic of knowledge. Oddly enough, the traditional direction of fit for our cognitive attitudes (mind-world) seems to be reversed in this case. It is not my purpose here to offer an epistemology of practical knowledge; just to remember that it is a kind of knowledge that, in a sense, wants to be true of the objects and facts known even if it is not controlled by what is known.

3. Epistemic privilege

In contrast with our knowledge of natural beings, we seem to have a better knowledge of artifacts insofar as they are products of our making. Maimon seemed to point to this feature with the analogy with the perfection of God's knowledge. Our philosophical tradition has been obsessed by the following idea: "we know better what we make ourselves". But what kind of epistemic privilege is here involved? And how is this privilege supported?

One possible explanation is given by the fact that the object has been *authored* by the makers. As I have said, we can have practical (intentional) knowledge of artifacts that would consist in this kind of knowledge that controls and regulates (in a normative and not merely in a causal way) the production of the artifact. This knowledge seems to be grounded on the fact that the maker is behaving as the *author* of the artifact in such a way that he or she knows it *because of that*.

This very fact could make sense of some of our intuitions about epistemic privilege in this case. Firstly, makers have the last word about what an artifact is (a notion of privilege); secondly, makers do not seem to observationally ground their knowledge of what they are making so that

maybe it is not grounded at all (a notion of groundlessness), and thirdly makers exhibit a certain presumption of truth about what the artifact is, in contrast with non-makers (a sort of authority).

Note that these features go beyond the mere fact that our knowledge of artifacts is immune to a limited range of error or that we are not universally ignorant about these things, as we can be about natural kinds (Thomasson 2003, 2007). On the one hand, it attempts to account for these features in terms of a distinctive kind of knowledge that makers possess, and that explains why culturally speaking we are not liable to certain errors or to universal ignorance. On the other hand, it recognizes the main role played by makers in the process of fixing the features that would characterize the artifacts themselves. In principle, it would be a way of developing an historical-intentional theory on the nature of artifacts that make stronger claims on the epistemological dimension.

Nevertheless, we lack yet some account of how the features associated with the putative epistemic privilege of makers are understood, and there are two possible lines of reasoning that we need to avoid.

4. How not to understand our epistemic privilege regarding our knowledge of artifacts

In this section, I will reject two different suggestions: the first one tries to derive epistemic privilege from a sort of prior access to those intentions that guide the making; the second one attempts to argue that being the author of the artifact is a sufficient ground for the kind of practical-intentional knowledge that she exhibits.

Let us consider first this last option. The idea is that we have a sort of knowledge by authorship of the artifact: by being the author of the object of knowledge (the artifact), the maker knows it. This way of addressing the issue raises an immediate worry: how is it possible that the mere fact of being the author *grounds* the truth of our knowledge about what we are making? In the case of intentional actions, where this model could have certain plausibility, it is the non-contingent link between the intentional action and the knowledge that seems to say something for this thesis. But this means that we need an independent support for the idea that the artifact to be known is constituted by the sort of knowledge in question. The contrast with the knowledge of our intentional actions becomes clear when we realize that intentions and other mental states are constitutive of the mental life of the subject in a sense in which artifacts cannot be. Even if we think of artifacts as mind-dependent objects, they entertain a certain metaphysical independence regarding the subjects, even if they are their makers.

The second option is less appealing because to make our knowledge of the artifact dependent on the knowledge of the intentions that guide its production is a way of severing the link

between the knowledge of our intentions and the knowledge we can acquire of the intentional actions and the product of their execution. We can immediately and authoritatively know what we are intending to do (to make an artifact), but that does not say anything about the knowledge of the artifact to be produced, that can be acquired and grounded independently of the making. Making a certain artifact depends on carrying out the intentions and the maker needs to know what it is to succeed and this cannot be given directly in the knowledge of the intentions.

5. A model for the epistemology of artifacts

If one is the maker of an artifact (the original maker or someone that makes new instances of the same artifact), then one knows what is to be an F and insofar as one makes an F through the realization of his or her intentions, one knows that it is an F that is being made. But the mere fact of being the maker does not ground the knowledge of the artifact; we need to preserve and explain the idea that this knowledge *fits* what is produced. I will propose to understand this *fit* in the following terms: the conception that guides the production of the artifact is sufficient to produce the artifact through a reliable mechanism that is under the control of the subject. And this is only possible if the subject is able to identify a set of success conditions learned within a dynamics of exchanges with an environment rich in functions and purposes. In so doing, the subject is able to fit her conception to the effective production of the artifact, *if she would perform the actions under her control*. In this way, I try to account for two intuitions: first, successfully making an artifact depends on a correctly informed plan the subject has access to; second, the knowledge does not have to precede the process of making itself, because this knowledge could be grounded in the process of making through which the subject acquires sufficient evidence about the conditions of success that are essential in effectively making the artifact in question. And this in a double sense: the subject understands what it is to be such kind of artifact; the subject is able to identify that, in such a case, she has been successful.

The maker achieves something of epistemological import in the making by virtue of exhibiting a conception that fits the artifact because it is sufficient to cause it to exist through a reliable mechanism under her control. At least three conditions need to be taken into account: first, maker's possession of a set of abilities; the successful production of the artifact that works according to a set of criteria; the grasp by the agent of the fact that the artifact exists because it is the result of the exercise of her own abilities.

Let us take a look at the conditions under which the agent may fail in her knowledge:

i) The object made does not work; so there is no success that could be ascribed to the maker.

ii) The agent does not understand that success is impossible to get given the abilities (and other necessary conditions) she has access to. The agent exhibits a bad understanding of what the relevant abilities involved in the making of an artifact are.
iii) The agent exhibits a mistaken conception of what the artifact is and consequently any attempt to make one would result in failure. This mistaken conception concerns the empirical conditions related to the good working of the artifact.
iv) The agent does not base her actions to make the artifact on a reliable mechanism; the production of the object is a pure coincidence or lucky.
v) The agent is not in a position to judge that the object exists in virtue of the conception she has of such kind of artifact and how this conception determines –in conjunction with the exercise of her abilities- the existence of the artifact as such artifact. This does not mean that the subject needs to explicitly judge in such terms in order to have the knowledge in question.

Many consequences derive from this set of possible features about the nature of the maker's knowledge. One of them can be put in the following terms: the authoritative nature of such knowledge is better explained when we understand the interdependence between the types of practical knowledge and inductive knowledge the maker can have of the artifact. If we know what an artifact is by being its author, this is because we ground our knowledge by controlling the production of it. Our practical knowledge cannot be severed from the mastery of information regarding the empirical conditions of what is under our control, in the same way that being receptive to this information depends on the practical anticipation of what we already know by being practically engaged in the process of controlled making of the artifact. Only this move could solve the difficulty of a notion of knowledge that seems to be far from our ordinary empirical assessments, that contravenes the idea that the object must regulate our cognitive attitudes and that reverts the direction of control that one takes to be characteristic of knowledge.

Acknowledgements
I would like to thank Diego Lawler, María Muñoz and Víctor Verdejo for comments on previous versions of this paper. This research has been funded by a research grant of the Spanish "Ministerio de Economía y Competitividad" (FFI2008-12054: *Epistemology of artifacts. Affordances, practical knowledge and epistemic artifacts*).

ARTIFACTS, EPISTEMIC PRIVILEGE AND CONCEPT POSSESSION

VÍCTOR M. VERDEJO
Universidad de Santiago de Compostela

ABSTRACT: Analysis of the precise relation between common views on concepts and artifacts is largely missing. In this paper, I address one central case of philosophical scrutiny: the relation between anti-individualism about concept possession and the epistemic privilege of authors as regards the artifacts they create. I show that these theses are in a formidable tension and that philosophers are well advised to choose one between the two. I argue, finally, that this choice leaves untouched the essentially intentional, mind-dependent nature of artifactuality.
KEYWORDS: possession conditions, privileged knowledge, anti-individualism, mind-dependence

Reflection on concepts is one of the central concerns of both philosophy and cognitive theorizing. From among the many areas of inquiry into the nature of concepts and thought a salient theme is the conditions for their possession. Even if universal agreement is hard to find in these territories, there is at least one thesis regarding concept possession that is largely shared among philosophers, namely, so-called anti-individualism (Burge 1979, 2007). For present purposes, we can state anti-individualism as in (A):

(A) The relations subjects bear to their (physical or social) environments are (at least partially) constitutive of the concepts they possess.

There are, I would say, a number of issues for which (A) involves a substantial revision or modification of pre-theoretical as well as philosophical views. These include self-knowledge of the content of one's own mental states, rationality or cognitive significance. In the context of this symposium, I wish to examine how (A), or theses of its anti-individualistic kin, affect our views on artifacts and artifactual kinds.

What are the connections between anti-individualism, understood as a theory of concept possession, and the theory of artifacts? It seems to me that there are at least two intersections that signal two crucial epistemological as well as metaphysical fields of study in this regard. One is the problem of the creation of new artifacts, or how new artifactual kinds are brought into existence – what Vega and Lawler (forthcoming) call the "creation requirement". The other, closely related issue, will be the main concern of this paper: the epistemic privilege of authors regarding the artifacts they create. In what follows, I will be defending that the intersection of the anti-individualistic theory of concepts and predominant views about the epistemic privilege of artifact

makers turns very much on a striking clash of intuitions and basic commitments. In particular, if anti-individualism is true, then it very much follows that there is no such a thing as the epistemic privilege of makers of artifacts.

This is a substantial result concerning the usual conception of artifacts. As several authors have pointed out (e.g. Hilpinen 2008), artifacts are, by definition, instances of a particular kind of object, namely, objects made with the intention to serve a given purpose. This definitional nature provides strong support for the claim that, unlike natural kinds, artifactual kinds are mind-dependent. By quite general assent indeed, creative intentions are necessary for genuine artifactuality (Hilpinen 1992; Bloom 1996; Vermaas and Houkes 2003). But if this is correct, note that it would seem that the putative creative intentions should be understood along the lines of epistemic privilege: authors have to know, in an infallible or authoritative way, what it is they are creating (who would otherwise!?). This is, in effect, a view explicitly defended by Amie Thomasson (2003, 2007).

There are many ways in which epistemic privilege for creators of artifacts may be stated. There is for instance an issue about whether the epistemic privilege in question concerns authors of objects belonging to an extant kind of artifact or else, authors of genuinely new artifacts. Another important question is what kind of knowledge epistemic privilege requires in this context. For instance, it is clear that epistemic privilege does not plausibly include second-order knowledge that one is the maker of an artifact or knowledge about the object's physical composition. Concessively, and for simplicity's sake, in the discussion to follow I will be focusing on what seems to me the most plausible version (E), for any author A, and object O.

(E) If O is a genuine artifact originally created by A, then A knows the sort of thing O is.

In this paper, I will defend that the relation between (E) and (A) above is one of unpalatable tension. Discussion will proceed in three steps. First, I will argue that, generally, (E) stands in direct opposition to (A) by promoting a very singular view about the especial relation between knowledge of an object O and the possession of its concept. Secondly, I will show how (A) can be used to argue for the plausibility of conceivable situations in which (E) is certainly false. These considerations run to the conclusion that (E) and (A) involve substantially different and incompatible views about the nature of concept possession, in such a way that theorists of concepts (of artifacts) are forced to choose one between the two. Finally, I show that decision between the defense of (A) and (E) has nothing whatever to do with mind-dependency because (A) is perfectly consistent with an intentional-dependency (as opposed to a knowledge-dependency) view of artifactuality.

1. Knowing and thinking

Perhaps one of the most salient features of contemporary discussion about concept possession is to distinguish sharply between knowledge conditions and thinking conditions. In a tradition that one may trace back to Frege, the guiding idea has been that, in many central cases, in order to possess a concept C of O one does not need to fully know O, not even if 'fully know' means something like 'fully satisfying some communal standard of knowledge'. This is, to be sure, one of the core tenets of anti-individualism (1979, 2007). This idea resonates in the work of authors that, while coming from different, indeed opposing traditions, would agree that strong conditions on knowledge are usually not (e.g. Peacocke 1998) or are never (e.g. Fodor 1998) adequate conditions on concept possession.

In this context, it should strike one as puzzling that, apart from stipulative cases, a thesis such as (E) is ultimately correct. The reason is that (E) amounts very much to a formidable knowledge condition on the possession of a given concept C of an artifactual object O. What (E) states is that for A to be a genuine creator of O, A must be incorrigibly right about the nature of O. Since there is no creation of O without creative intentions regarding O, and since there are no creative intentions of O without the corresponding concept of O, (E) entails that A has to have privilege knowledge of O in order to possess its concept C.[4]

Note that this clash of views regarding concept possession of (A) and (E) will concern, generally, the very same sort of artifactual objects. Patently, anti-individualist analyses of artifactual cases are legion. Only in Burge (1979), one finds examples involving sofas, clavichords or contracts. This is because artifacts provide very good examples of the anti-individualistic assumption that beliefs infected by partial understanding are "rampant in our pluralistic age" (Burge 1979, 107). Note further that the epistemological role of makers with respect to their creations is analogous to the role of experts with respect to their field of expertise. However, unlike makers, experts are not, on standard anti-individualistic views, plausibly required to have epistemic privilege about the nature of their subject matter (think what it would take for economists to have this knowledge!).

In sum, it would seem that one cannot have it both ways. From a general point of view, acceptance of either (A) or (E) yields theories of concept possession that blatantly pull in opposite epistemic directions.

2. Resurrected Caroline

[4] Note further that, since there are no artifacts without creators of them, and correspondingly no concepts of artifacts without their creators, (E) involves a universal condition on possession of a concept of O: if there exists, as it were, a possessor of the concept of O, then there exists or existed a creator A with privilege knowledge about O's nature.

The compatibility between (A) and (E) is further put to the test by the conceivability of certain anti-individualist scenarios. These scenarios make amply clear that anti-individualism is a theory especially demanding in the discussion about artifacts. To illustrate, let me briefly introduce the case of Resurrected Caroline.[5]

Caroline is a perfectly rational thinker in a community like our own except for the fact that there are no sofas. We can imagine Caroline at some point decidedly engaging in the project of creating a new kind of object, a piece of furniture meant for sitting. After months of hard work, Caroline succeeds in her project and brings into existence an object that we would, in actual communities, undoubtedly call "sofa". We may imagine further that Caroline decides to actually call her invention "sofa". Unexpectedly, the creation process severely undermines Caroline's health and she is about to die before she can advertise her creation to any other members of her community. Before dying, she writes and puts on her creation the note: "This is a sofa. It is made or meant for sitting".

Now, after her death, sophisticated and brave thinkers are the ones first acquainted with Caroline's creation. Intensive discussion follows. In the end, in spite of her note, it is decided that Caroline's "sofa" is not really made or meant for sitting. On the contrary, members of Caroline's community judge that the thing they call "sofa" is a work of art or religious artifact. In a new sort of worship, "sofas" are reproduced and exhibited all over the place. The key point is that in this scenario, and even if people use the form of words "sofa" to refer to Caroline's creation, there are no sofas. Many of the subsequent instances of "sofa" would collapse under a person's weight. Let us call "safos" the relevant objects. If (A) holds, it follows that, in this scenario, there are no sofas but there are lots of safos instead. It follows from this that Caroline did not create a sofa at all but the first paradigm instance of a safo.

This scenario introduces a perfect counterexample to (E). What we have said so far suffices to make plausible the idea that Caroline lacked any sort of epistemic privilege about the thing she created. Indeed, the scenario shows that, if (A) is true, Caroline was wrong, and even substantially wrong, about her invention, that is, about safos: she thought that her safo –the thing she called "sofa"– were a furnishing kind of thing whereas, as it happens, it is a religious kind of thing.

Close examination of this argument is beyond the scope of this paper. However, I would like to present a continuation of Caroline's story that makes its moral still more vivid: let us imagine that, after some years, Caroline comes back to life. The details of why this is so need not concern us

[5] This case summarizes a variant of the scenario presented in Verdejo (manuscript) which in turn is inspired by a case originally introduced by Burge (1986).

here (more veridical construals are ready to hand). Caroline can in this case look and see what happened to her creation. Would not she insist on the appropriateness of her original conception?

Another way of putting this question is as follows. For any subject S, creation X, and environment E, how would a rational S react in the light of unexpected information about X and E? As Schroeter has argued at length (Schroeter 2008), the answer to this kind of question is more grist to the anti-individualist mill. In particular, when Caroline comes back to life, she is rationally required to refine her own representational practices regarding the X she created in the light of the new and highly surprising information about her environment. It is hard to see how Caroline could rationally dispute the claim that, in spite of what she believed and knew, she was the first maker of a safo-sort of thing even if she thought, wrongly, that safos were meant for sitting.[6]

In short, (A)-analyses of easily conceivable scenarios involve a direct denial of (E). This further shows that one of them, either (A) or (E), will have to go.

3. Two kinds of mind-dependency

The final line of argument that I wish to present in this paper is devoted to show that both (A) and (E) are compatible with the intuitively plausible and standard view that artifacts are mind-dependent. As noted above, the baseline definition of artifact appeals to a process of creation of an object for certain, intended purposes (e.g. Hilpinen 2008). Roughly then, since one does not find intentions without minds, it is difficult to deny that, as standardly understood, artifacts are mind-dependent: they depend on the putative intentions of their creators.

In this context, it is fundamental to arrest a certain temptation that the foregoing considerations may encourage. This is the temptation to see the choice between (A) and (E) as the choice between, say, an alienated or otherwise inhuman conception, and an intentionally articulated conception of artifactuality. Nothing of the sort really happens. For this reason, mind-dependency does not force on us the preference for an (E)-epistemic view about the concepts of artifact creators regarding the artifacts they create. Anti-individualism is utterly compatible with mind-dependency.

All that mind-dependency requires, in the present context, is that for any author A of an object O, A has the appropriate intention I, where I is an intention of creating O –call it the intention I(O). Clearly, the only way for A to have an intention of the required I(O)-sort is for A to possess the concept C of O. But all this is perfectly compatible with anti-individualism. Anti-individualism poses a substantial constraint on the individuation of the concepts a subject possesses, namely, that they be partially so individuated by the physical or social environment. It follows that, a fortiori, it

[6] Compare with the classical example: after knowing about his mistake, the medically ignorant subject could not rationally dispute the claim that he thought, wrongly, that arthritis is also a disease of thighs (as opposed to thinking, correctly, that tharthritis is also a disease of thighs).

poses also a substantial constraint on the individuation of artifacts authors create. But it does not follow that it allows for the possibility of mind-independent alienated artifacts, that is, cases in which an author A of O lacks the concept of O or the corresponding intention I(O). It's only that As might not know, let alone know in a privileged way, the nature of the thing they are creating or the correct way to individuate the concepts and thoughts they employ in its creation.[7]

What this suggests is, of course, that we should very well distinguish between two dependencies that are, who would care to deny, clearly distinct: intentional-dependency and privilege-knowledge dependency. The latter kind of dependency is much stronger and is the one naturally associated with (E). The choice between (E) and (A) is therefore more precisely the choice between these quite different kinds of mind-dependency.

Conclusion

To be sure, the relation between the theory of concepts and the theory of artifacts awaits further exploration. In this paper, I have set myself the task of laying out an unresolved and unacceptable tension regarding substantial views on concept possession and makers knowledge. If the view here presented is sound, our forthcoming full theory of concepts (of artifacts) will sooner or later have to choose between a theory that respects anti-individualistic claims and one that satisfies our epistemological intuitions about artifact creation.

In the meantime I would like to put my bets very much on the former possibility. For one thing, it seems fair to consider anti-individualism as the mainstream view on concepts and attitude ascription, including of course *artifactual* concepts and attitude ascription. For another thing: I would not be in the least surprised if anti-individualism turns out to be our best and most precise articulation of the view, stressed by champions of the theory of artifacts, that many artifacts have a fundamental public dimension (Dipert 1993; Thomasson forthcoming), one according to which they "do not depend on the *individual* intentions of their makers; they also depend on public norms" (Thomasson forthcoming, 59, emphasis hers).

[7] Note that this is something orthogonal to issues about self-knowledge of thought and conceptual contents. Supporters of (A) can agree, and arguably do agree (see Burge 1988), that authors always know the contents of their C of O. What they might certainly not know (let alone know in a privileged way) is the nature of the Os involved and, precisely because of this, the proper individuation of the concept C of O. This has nothing particular to do with artifactual concepts: I may be infallibly right about the content of my *water*-thoughts and have no knowledge, let alone a privileged knowledge, about the (chemical) nature of the substance or the proper individuation of my concept *water*.

Acknowledgements:

This research has been partially funded by the Ministerio de Economía y Competitividad, Spanish government, under the research project FFI2009-08828/FISO, and the Catalan government, via the consolidated research group GRECC, SGR2009-1528.

References

Anscombe, G.E.M. 1957, *Intention*, Oxford: Blackwell.

Barsalou, L.W. 1987, 'The instability of graded structure: implications for the nature of concepts', in Neisser, U. (ed.), *Concepts and Conceptual Development*, Cambridge, Cambridge University Press.

Basalla, G. 1988, *The Evolution of Technology*, Cambridge, Cambridge University Press.

Bloom, P. 1996, 'Intention, history, and artifact concepts', *Cognition* 60, 1-29.

Burge, T. 1979, 'Individualism and the mental', reprinted in Burge, T. (ed.) *Foundations of Mind*, Oxford, OUP, 2007, 100-50.

Burge, T. 1986, 'Intellectual norms and the foundations of mind", *Journal of Philosophy* 83, 697-720.

Burge, T. 2007, 'Postcript to "Individualism and the mental"', in Burge, T. (ed.), *Foundations of Mind*, Oxford, OUP, 2007, 151-191.

Costall, A. 1995, 'Socializing affordances', *Theory and Psychology* 5, 467-481.

Dipert, R. 1993, *Artifacts. Art Works and Agency*, Philadelphia, Temple University Press.

Fodor, J.A. 1998, *Concepts. Where Cognitive Science Went Wrong*. Oxford, OUP.

Gibson, J.J. 1979, *The Ecological Approach to Visual Perception*, Boston, Hoghton Mifflin.

Hampshire, S. 1975, *Freedom of the individual*, London: Chatto & Windus.

Heidegger, M. 1967, *Being and Time*, Cambridge, Blackwell.

Hilpinen, R. 1992, 'On artifacts and works of art". *Theoria* 58, 58-82.

Hilpinen, R. 2008, 'Artifact', in Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/fall2008/entries/artifact/

Kornblith, H. 1980, 'Referring to artifacts', *The Philosophical Review* 89, 109-14.

Maimon 1904, *Guide for the Perplexed*, Friedländer translation, London: Routledge & Kegan Paul.

Malt, B. C. and Sloman, S. A. 2007, 'Artifact Categorization: the Good, the Bad and the Ugly', in Margolis, E. and Laurence, S. (eds.), C*reations of the Mind*, Oxford, Oxford University Press.

Neisser, U. 1987, 'From direct perception to conceptual structure', in Neisser, U. (ed.), *Concepts and Conceptual Development*, Cambridge, Cambridge University Press.

Nelson, J. 1982, 'Schwartz on reference', *The Southern Journal of Philosophy* 20, 359-65.

Putnam, H. 1975, 'The meaning of "meaning"', in Gunderson, K. (ed.), *Language, Mind and Knowledge*, Minneapolis, University of Minnesota Press, 131-93.

Peacocke, C. 1998, 'Implicit conceptions, understanding and rationality', *Philosophical Issues* 9, 43-88.

Preston, B. 2006, 'Social context and artefacts function", *Studies in History and Philosophy of Science* 37, 37-41.

Schroeter, L. 2008, 'Why be an anti-individualist?', *Philosophy and Phenomenological Research* 74, 105-41.

Schwartz, S. P. 1978, 'Putnam on artifacts', *The Philosophical Review* 87, 566-74.

Schwartz, S. P. 1983, 'Reply to Kornblith and Nelson', *The Southern Journal of Philosophy* 21, 475-79.

Thomasson, A.L. 2003, 'Realism and human kinds', *Philosophy and Phenomenological Research* 67, 580-609.

Thomasson, A.L. 2007, 'Artifacts and human concepts', in Margolis, E. and Laurence, S (eds.), *Creations of the Mind: Theories of Artifacts and their Representation*, Oxford, OUP, 52-73.

Thomasson, A.L. forthcoming, 'Public artifacts, intentions, and norms', in Franssen, M., Kroes P., Thomas, A.C.R. and Vermaas P.E. (eds.), *The Metaphysics of Technical Artifacts*, Synthese Library.

Tomasello, M. 1999, 'The cultural ecology of young children's interactions with objects and artifacts', in Winograd, E. Fivush, R. and Hirst, W. (eds.), *Ecological Approaches to Perception*, Mahwah, Lawrence Erlbaum Associates.

Vega, J. and Lawler, D. forthcoming, 'Creating artefactual kinds', in Franssen, M., Kroes P., Thomas, A.C.R. and Vermaas P.E. (eds.) *The Metaphysics of Technical Artifacts*, Synthese Library.

Verdejo, V.M. manuscript, 'Against artefactual epistemic privilege'.

Vermaas, P. and Houkes, W. 2003, 'Ascribing functions to technical artefacts: A challenge to etiological accounts of functions', *British Journal for the Philosophy of Science* 54, 261-89.

PHILOSOPHY AND METODOLOGY OF SCIENCE

FILOSOFÍA Y METODOLOGÍA DE LA CIENCIA

MECANISMOS Y MODELOS EN BIOLOGÍA Y EN CIENCIAS COGNITIVAS

Coordinador Antonio Diéguez
Universidad de Málaga

La filosofía de la ciencia mas reciente, en especial el enfoque semántico, subraya la importancia de los modelos en la práctica científica, así como en la explicación y comprensión de los fenómenos de los que se ocupa la ciencia. Tanto en el ámbito de las ciencias biomédicas como en el de las ciencias cognitivas se ha prestado una atención especial a los modelos mecanicistas, que muchos autores consideran como una alternativa al modelo más clásico de explicación nomológica (Bechtel & Abrahamsen 2005). Para estos autores los mecanismos son estructuras que realizan una función en virtud de sus partes componentes, sus operaciones y su organización, de forma que el funcionamiento del mecanismo es responsable de un fenómeno.

El propósito de este simposio es analizar y discutir los modelos en biología y en ciencias cognitivas, planteando varios asuntos filosóficos en torno a las peculiaridades de los modelos mecanicistas. La función representativa y explicativa que los modelos desempeñan en la biología y en las ciencias cognitivas es cada vez más discutida en laliteratura, pero quedan aún muchos aspectos relevantes por dilucidar. Hayuna cierta tendencia a reducir esta función de los modelos a la aplicación de mecanismos, pero, si bien los mecanismos ocupan un lugar preeminente enla explicación mediante modelos, el análisis mecanicista no agota todaslas formas en que estos son empleados para proporcionar explicaciones.

Las contribuciones que proponemos plantean cuestiones variadas dentro de este marco, algunos de los temas y cuestiones que se plantean son los siguientes:

1. Si la noción de organización biológica entendida como el funcionamiento de un mecanismo puede al mismo tiempo garantizar la comprensión de las enfermedades como el mal funcionamiento de ciertos mecanismos.

2. Cómo pueden conjugarse o integrarse en el estudio de la complejidad biológica la metodología mecanicista basada en el análisis de partes funcionales y la reconstrucción de comportamientos, y la modelización predictiva, basada en el descubrimiento de formas de regularidad (leyes o principios generales).
3. Cual es la contribución del enfoque mecanicista al estudio de los procesos mentales y la repercusión que ello tiene en la consideración de los mismos como procesos corporeizados.
4. Cual es el papel explicativo de los modelos científicos en relación al enfoque más tradicional de explicación científica, tal como la explicación por cobertura legal, y si puede argumentarse que el fin último de los modelos es proporcionar comprensión.
5. Si el concepto de nivel de organización que se maneja en el enfoque mecanicista es útil para caracterizar la noción de organización biológica y cuáles son las consecuencias del uso de una noción apropiada de nivel para las discusiones filosóficas sobre el reduccionismo o la causación descendente.

En suma, la sesión permitirá plantear varias cuestiones específicas que relacionan problemas clásicos de la filosofía de la ciencia general con cuestiones filosóficas específicas que surgen en las áreas de la biología, la ciencia cognitiva o la medicina. Todas ellas tienen en común la atención prestada a la epistemología mecanicista y la forma en la que se encara las dificultades que plantea la complejidad de la organización biológica y la explicación científica de los fenómenos que derivan de ella.

REFERENCIAS

BECHTEL, W. & ABRAHAMSEN, A. (2005). “Explanation: A mechanist alternative”. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421–441.

MECHANISMS IN BIOLOGY AND MEDICINE: BIOLOGICAL ORGANIZATION AND PATHOLOGIES

ARANTZA ETXEBERRIA
Department of Logic and Philosophy of Science
Euskal Herriko Unibertsitatea UPV/EHU
Donostia – San Sebastián, Spain
arantza.etxeberria@ehu.es

Introduction

In a recent paper Nervi (2010) has argued that malfunction should not be understood as "a mirror image of function" (216), because the study of pathologies does not necessarily arise as knowledge of how physiological mechanisms are impaired. In his view, pathology or malfunction and physiology are "independent" from each other. Following to a certain extent Canguilhem's framework to understand health and disease (1966/1991), Nervi claims that we cannot assume that a pathological fact is the negation of a physiological one, because medicine aims to find therapies. Thus, pathologies need to be pragmatically understood in the context of the need to find therapies or cures, and the procedures adequate to that end may oblige to depart from the context of the physiological knowledge.

This problem affects our ideas about how medicine needs to understand living organization. Nervi´s claim suggests that pathologies have an independent organizational logic or at least that medical knowledge does not rely on the knowledge we have of the nature of positive inputs to biological organization. The biological organization of an individual organism has been conceptualised in terms of how its parts operate so as to contribute to the maintenance of the whole. In this framework one would think that a disease will have to do with some contribution not being accomplished, either because the part is damaged, because it is prevented from fulfilling its role by some internal or environmental cause or even because it was an evolutionary adaptation that is not adaptive any more. But it is difficult to see how pathological phenomena may not be failures to contribute.

Nervi's arguments for his case are basically pragmatic; he offers examples in which he shows that sometimes the discovery and research of therapies needs to follow the logic of the pathology and not of the supposed physiological function that has been perturbed. But he also mentions questions of principle as he argues that there is an underlying assumption that knowledge of the physiology of the system should be valuable to negatively inform about the disease (what is prevented) and positively about its cure (how to repair or regenerate the contribution).

The notion of disease is evaluative in the sense that to be a disease a condition needs to be considered negative or bad (Murphy 2008). There are no diseases in the natural world until someone considers that a given condition is bad, then a description of a condition by itself would not be a disease. Yet different agents can carry out the negative evaluation involved: the patient herself, also medical professionals or even social agents, such as members of the family (Casado da Rocha and Etxeberria 2012) in multidimensionality of values.

2. Pathological mechanisms and physiological mechanisms

The discussion started by Nervi's paper is set within the mechanist approach to explanations of organic phenomena. According to Bechtel and Abrahamsen (2005, 423) "A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated mechanism is responsible for one or more phenomena". This approach is considered to be as useful to understand pathological phenomena as physiological ones, insofar as both can be described as mechanisms. In this, Nervi follows Craver (2001) who explicitly maintains that his account of mechanistic role functions "does not appeal to any sense of adaptiveness in an environment; instead it appeals only to roles in contextual systems [...which] may be adaptive or destructive." (67). In this sense, they follow Cummins's descriptive approach to mechanistic role functions, as devoid of appeals to natural or intrinsic normativity. "The assignment of a role is enough of a criterion to define the mechanism, independently of whether the mechanism be overall adaptive or maladaptive" (217). Thus, both Craver and Nervi consider that mechanistic explanations have no commitment with evaluations of the utility or the validity of the proposed mechanisms. For Nervi the mechanism of the malfunction may be independent and different from the malfunction of a physiological mechanism.

This view of mechanisms as purely conventional and devoid of the normativity to distinguish the pathological from the physiological lies behind Nervi's criticisms to the idea that "there is an implicit agreement about the fact that a mechanism must be valuable for the organism" (217) and that therefore "malfunction is conceptualized as a failure in one or more steps in the physiological sequence of events". Although Nervi is ready to accept that this approach is not per se

inconsistent and can even be usefully adopted, according to him, what is or can be sometimes pragmatically wise cannot be generalized for all medical knowledge. Thus he maintains that pathological mechanisms are independent when the natural history of the disease has primary importance and it is necessary to describe "causal chains of pathological events that lead from the initial aetiology (if known) to the possible outcomes of that particular disease" (218). For example these causal chains are taken into account when describing different kinds of diabetes so that therapies can interrupt them at the best possible point. In addition pathological phenomena affect sets of organs that are not considered to be physiological systems.

Then medicine needs to understand the mechanisms implied in malfunction to search for adequate therapies, and according to Nervi pathologies have sometimes an independent nature that does not allow understanding them as malfunctioning physiological mechanisms. Then they must be understood as independent entities. Moghaddam-Taaheri (2011) has analysed the problem raised by Nervi as a discussion about whether diseases can be seen as broken mechanisms. In her view conceptions of pathologies as mechanisms that are broken is more practical and relevant to find therapies, and argues that in fact that is the usual procedure used when developing drugs On the other part, Gross (2011) considers that there are principled reasons funded in systems biology to adopt what he calls the "autonomous entity" (Nervi's "independent entity") approach, because there disease appear as "caused by network perturbations and might correspond to network states that themselves exhibit organization and robustness" (490-91). For Marcum (2011) Nervi's stance is relevant to consider whether there is progress in medicine in terms of whether there is cure, and how can we define what cure is, whether to restore some capacities can count as a cure or only if the primitive physiological mechanism is back. The example used is type 1 diabetes, a form of diabetes that is normally treated by injecting insulin to the patient whose pancreatic cells are unable to secrete this hormone. Although the treatment dramatically saves the life of patients and provides a relative quality of life, it can be discussed whether it constitutes a cure of the disease or whether only a complete restoration of the patient to the state without the disease can be considered as such.

In sum, one kind of discussion triggered by Nervi's paper concerns the pragmatic goal of looking for therapies and the broken/normal view or independent entity heuristics involved in that endeavour. In the rest of the paper I will turn to the one concerning biological organization.

3. Biological organization, mechanisms

The role of biological organization is to account for how entities and their activities produce a phenomenon of interest as a contribution to the phenomenon. However, to find out which particular form of organization lies behind a particular phenomenon is part of the project and it is

considered to be a matter of empirical discovery in the mechanist approach (Illari & Williamson 2010). Complex systems are characterized by their organization, by the relations among components such as spatiotemporal, feedback and control, self-organization and constraint, emergence or downward causation are among the relations taken into account to account for the organization of a system. Some of these relations have also been studied in formal mathematical terms within Systems Biology. When the notion of organization is associated with complexity and holistic systemic properties the mechanistic framework with its emphasis in decomposition seems to be insufficient.

Thus when considering the relation of biological organization and mechanisms there are two different approaches to the notion of organization in current biology and philosophy, a weak one and a strong one. Whereas the weak one is a generic notion of organization that attends to the arrangement of component parts and processes in the target system, the strong one conceives that the organization is the clue to define very complex phenomena such as life and health, and it crucially includes a teleological normative dimension (such as the capacity to self-produce or self-maintain) as the phenomenon the organization needs to grasp.

Now, to approach Nervi's challenge that pathological entities need to be considered by Medicine independently of the physiological states and mechanisms that characterise health it is necessary to consider discussions in the philosophy of medicine about naturalist or normative conceptions of diseases or pathologies. The philosophy of medicine debates about the nature of health and disease, especially about whether diseases can be described appealing to science or to biological theories, thus providing descriptions which do not involve norms and evaluations (which normativists consider inevitable). In naturalist terms the question would be if biology can provide a theory able to delimit diseases. This question obliges to think also on how the evaluations that most consider inevitable for the notion of disease or pathology can be done.

Three different positions toward normativity may be used to distinguish three different approaches to how biological theories can be normative to evaluate that a condition is pathological. NN1 maintains that there is no natural normativity; NN2 considers that different theories based on different concepts of biological function and dysfunction can be pragmatically used to delineate and define diseases, as each of them uses a different intuition about normativity; and, finally, NN3 is the view that organisms have a natural (living) normativity, associated with the strong notion of organization reviewed in the previous section. Problems of space oblige to keep this very brief, a longer version will be available soon.

NN1 = no natural normativity

NN1 maintains that biological theories are completely neutral with respect to the valuable or destructive nature of the mechanisms studied, and hence they cannot define what is "natural" or healthy, or what is pathological. For Ereshefsky (2009), who is examining Boorse's account of disease as malfunction (1977), this means that biology cannot define pathological states in a neutral way, as it can only provide descriptions that need to be interpreted under the adequate circumstances. His main arguments to say that biology cannot provide a theory to delimit natural from non natural states are argument of variation within the human species, and the argument that descriptions are idealizations. However, if one maintains this the question arises: where do medical practitioners find the evidences to produce the normative claims affecting judgments about health and disease?

Philosophers like Grene (1976) have considered that descriptions can be normative to a certain point, and so do other essentialist or realist philosophers who consider that an adequate description/explanation of an anomaly or a disorder, such as those that appear in medical textbooks can and do very often have a normative role and be an aid for the medical practitioner that is trying to classify a condition as disease.

NN2 = naturalist normativity

In the philosophy of medicine, since Boorse, naturalist accounts consider that diseases are malfunctions, that is to say deviations of the normal functioning of physiological parts and processes that may be accounted for by the natural sciences. Although Boorse considers these accounts to be value neutral many replies have understood that a particular form of normativity, that which values health as a statistical normality is in play in Boorse's account. In fact, Fulford (2002) has convincingly argued that any account based on our knowledge of functions to analyze pathologies as dysfunctions is intrinsically normative.

The main notions of naturalistic normativity being handled by the literature are evolutionary. Two of them rely in evolutionary theory, one being a form of Wright´s etiological account of function, in which the contribution of the item is to to evolve by natural selection and evaluated according to NS in the past. The other is Boorse's notion according to which the contribution is to survival and reproduction, and is evaluated according to statistical normality. Different notions of function developed in the literature can ground normatively judgments about pathologies, and thus it is possible to say that there are different kinds of normativity being applied by the different conceptions of biological function.

NN3 = natural (living) normativity

Finally, some perspectives maintain that any organism is intrinsically normative, because to live means always that some situations are going to be preferred rather that others. This is

associated to the notion of strong organization that was presented previously. Canguilhem quotes Bichat saying that there cannot be a subject matter of physics called pathology because there is no notion of intrinsic value in physics or chemistry.

Yet, the notion of a natural normativity does not mean that it is transparent and easily achievable by the organism or person itself or by the medical practitioner. Canguilhem famously repeated Leriche's sentence that health is "life lived in the silence of the organs", and "disease is what irritates men in the normal course of their lives and work, and above all, what makes them suffer" (1991, 91). However the norm that characterizes health is only known when it is lost, whereas to recover is not necessarily to return to the initial situation but to establish the same or another, transformed, norm.

In addition, the strong notion of biological organization is one of perfection or optimality. Instead we should consider health as a state of dynamically and continuously redefining the organism's own norms under the thread of precarious circumstances. Precariousness comes in degrees, this gradation may go from mild difficulties to life-threatening conditions, and that may make it easier to accommodate illness and disease within health.

However, not all scientific work about pathologies follows a functional scheme neither is purely descriptive. Some models based on biological organization in different levels (tissues, cells, genes, etc.) attempt to take into account the complexity of organisms from a systemic perspective that tries to overcome weak organization and approach stronger forms of organization from a systemic approach.

In sum: health or the norm is not necessarily the absence of disease but a positive capability that is modulated according to the different degrees of precariousness caused by multiple sources, including illnesses and diseases. Thus, from the point of view of NN1, there is nothing that makes pathological mechanisms the opposite of physiological ones, both can be considered to be independent of each other. From the point of view of NN2, pathologies will nearly always be considered to be the opposite of physiological functions as they are evaluated as malfunctions according to the normativity implied by the function being appealed. From the point of view of NN3 pathologies are bad, but the normal physiological condition is internal and its knowledge is not.

Conclusion

This paper examines if the notion of disease can rest only in the understanding of function as a contribution to the organization of the individual organism. If not, some of the consequences can be that a) medicine makes pragmatic use of diverse conceptualisations of malfunction, some of

which can not be captured as the negative of function; b) different kinds of normativity operate in medicine to account for malfunction and disease (which are always considered to be normative); and 3) the notion of biological organization may be reconceived and extended so as to integrate the occurrence of pathologies in various degrees (this affects the notion of health and disease).

Acknowledgements

Funding for this research was provided by the Subvención a Grupos IT 505-10 of the Basque Government-Eusko Jaurlaritza, and Projects FFI2011-25665 and FFU2009-12895-CO2-02 of the Ministerio de Ciencia e Innovación.

REFERENCES

Bechtel, W. and Abrahamsen, A. 2005, 'Explanation: A Mechanistic Alternative', *Studies in History and Philosophy of the Biological and Biomedical Sciences , 36, 421-441.*

Boorse, C. 1977, Health as a Theoretical Concept, *Philosophy of Science* 44: 542–573.

Canguilhem G. 1991 *The Normal and the Pathological*. New York: Zone Books. (original en francés de 1966, Paris: PUF).

Casado da Rocha, A. and Etxeberria A., 2012, 'Towards autonomy within illness, applying the triadic approach to the principles of bioethics'. In H. Carel & R. Cooper Eds *Contemporary Topics in Philosophy of Medicine,* Newcastle: Acumen, in press.

Craver, C. 2001, 'Role functions, mechanisms and hierarchy', *Philosophy of Science* 68: 53-74.

Ehreshefsky, M. 2009, 'Defining "health" and "disease"'. *Studies in the history and Philosophy of Biology and Biomedical Sciences* 40: 221-227.

Fulford, K.W.M. 2000, 'Teleology without tears', *Philosophy, Psychiatry and Psychology* 7, 77-94.

Grene, M.1976, 'Philosophy of Medicine: Prolegomena to a Philosophy of Science'. In *PSA Proceedings of the Biennal Meeting of the Philosophy of Science Association*, 2, 77-93.

Gross, F. 2011, 'What systems biology can tell us about disease', *History and Philosophy of the Life Sciences*, 33(4): 477-496.

Illari, P. M. & Williamson, J. 2010, 'Function and organization: comparing the mechanisms of protein synthesis and natural selection', *Studies in History and Philosophy of Biological and Biomedical Sciences* 41, 279-291.

Jacob, F. 1993, *The logic of life*, New York, Pantheon Books.

Machamer, P, Draden, L & Craver, C. (2000) Thinking about mecanisms, *Philosophy of Science* 67 (1): 1-25.

Marcum, J, 2011, 'Medical cure and progress', *Perspectives in Biology and Medicine* 54, 176-188.

Moghaddam-Taaheri, S., 2011, 'Understanding Pathology in the Context of Physiological Mechanisms: The Practicality of a Broken-Normal View', *Biology and Philosophy* 26 (4):603-611.

Moss, L., 2012, 'Is the philosophy of mechanism philosophy enough?', *Studies in the History and Philosophy of Biological and Biomedical Sciences* 43, 164-172.

Murphy, D. 2008. Health and disease. In A. Plutynski & S. Sarkar (Eds.), *The Blackwell Companion to the Philosophy of Biology* (pp. 287-297), Balckwell.

Nervi, M 2010, 'Mechanisms, malfunctions and explanation in medicine', *Biology and Philosophy* 25: 215-228.

EL PAPEL DE LOS MECANISMOS EN LOS SISTEMAS BIOLÓGICOS

ÁLVARO MORENO
Universidad del País Vasco

Introducción

La complejidad de los sistemas vivientes es muy diferente que la de los sistemas no-vivientes, pues implica diversidad funcional, organización en niveles, estructura en diferentes redes imbricadas, sostenibilidad a largo término e historia acumulativa y capacidad agencial adaptativa. La evolución biológica muestra que la aparición de mayor complejidad va de la mano de la invención de jerarquías de control funcional de la misma. Como dice J. Mattick (2004), "lo que realmente importa en la evolución no es la generación de complejidad, sino su control selectivo y funcional". Lo que ocurre en los sistemas vivientes es un funcionamiento en diferentes velocidades, generándose dinámicas desacopladas, las cuales permiten que una parte del sistema explore un espacio de funcionalidades libre de las constricciones que supone mantener los procesos básicos. Cuando el sistema desacoplado superior encuentra una serie de soluciones funcionales, se retienen, convirtiéndose así en mecanismos de control sobre el sistema inferior. A su vez, el nivel "inferior" es fundamental, porque constituye la base del sistema, incluyendo la condición de posibilidad del sistema "superior" o regulatorio.

Pero la posibilidad de que aparezcan y se mantengan sistemas complejos depende de la existencia de mecanismos de control. Así, en la organización de los seres vivos se combinan procesos holísticos y partes distinguibles que pueden ser descritas como mecanismos. Esta combinación puede darse a varios niveles, y generar nuevas formas de holismo, pues las interacciones entre partes funcionalmente distinguibles a un nivel pueden generar a un nivel superior fenómenos emergentes globales. En definitiva, lo que caracteriza al funcionamiento de los seres vivos es una combinación compleja e internivélica entre procesos mecanísticos y holísticos. Lo podríamos llamar un "holismo de segundo grado", surgido de la complementariedad entre

mecanismos funcionales y procesos masivamente distribuidos (Moreno et al. 2011). Vamos pues en esta comunicación a desarrollar esta idea.

1. Mecanismos en el marco biológico.

La profunda circularidad organizacional, así como la combinación de relaciones causales internivélicas hace que debamos repensar el concepto de mecanismo en los sistemas biológicos. Recientemente, Bechtel y Abrahamson (2011) han reconocido este hecho señalando que

> Oscillations and other complex dynamical behavior can emerge when operations in mechanisms are organized non-sequentially, the interactions are non-linear, and the system is partially open to energy; moreover, such dynamic behavior may play an important role in regulating other operations in the mechanism. Understanding such dynamic behavior of mechanisms typically requires the tools of computational modeling or dynamical systems theory. Sometimes this modeling is done in isolation from mechanistic decomposition of the system into parts and operations. When these two kinds of analysis are coordinated, though, some or all of the variables and terms in the mathematical model correspond to properties of identified parts and operations (e.g., the rate of an operation). Bechtel and Abrahamsen call the accounts that result from analyzing the orchestration of operations in this manner dynamic mechanistic explanations.

Tomemos un ejemplo dentro de la organización celular, el de la ATPasa. Este agregado macromolecular situado en la membrana, con sus piezas ensambladas en una estructura bien determinada, cuando recibe un input de energía, realiza la tarea de bombear iones de sodio-potasio. Podemos ver que la identificación de partes con roles funcionales bien específicos, formando una estructura de relaciones causales secuenciales debe combinarse. El funcionamiento se produce por un cambio de conformación en la proteína cuando es fosforilada por el ATP. Como el resultado de la catálisis es el movimiento transmembrana de cationes, y se consume energía en forma de ATP, su función se denomina transporte activo, es decir, realiza el transporte contra gradiente, gastándo para ello energía. La demanda energética es cubierta por la molécula de ATP, que al ser hidrolizada, separa un grupo fosfato, generando ADP y liberando la energía necesaria para la actividad enzimática. Los procesos sucesivos del mecanismo son: 1) Unión de tres Na+ a sus sitios activos. 2) Fosforilación que induce a un cambio de conformación en la proteína. Esta fosforilación se produce por la transferencia del grupo terminal del ATP a un residuo de ácido aspártico de la proteína. 3) El cambio de conformación hace que el Na+ sea liberado al exterior. 4) Una vez liberado el Na+, se unen dos iones de K+ a sus respectivos sitios de unión de la cara extracelular de las proteínas. La proteína se desfosforila produciéndose un cambio conformacional de ésta, lo que produce una transferencia de los iones de K+ al citosol. La acción de este mecanismo es funcionalmente

relevante pues contribuye muy fundamentalmente al mantenimiento de la organización celular, al evitar la crisis osmótica y el consiguiente estallido de la célula.

Por tanto, la ATPasa o "bomba de iones" es un mecanismo en sentido pleno: esta formada por un conjunto de entidades y actividades cuya organización genera una serie de cambios regulares, caracterizables como una sucesión de procesos con un comienzo y una terminación bien definidos. Pero este mecanismo, formado por un conjunto de estructuras articuladas funcionalmente, depende a su vez del mantenimiento de otros mecanismos celulares: la enzima se degrada y ha de ser repuesta; otros muchos procesos y componentes son necesarios, de forma que en conjunto unos dependen de otros y en definitiva, el sistema se mantiene tanto por la contribución específica de cada mecanismo como éstos a su vez por su pertenencia a una organización disipativa (el conjunto de reacciones que constituyen el metabolismo celular) y en muy buena medida, distribuida. Es este contexto reticular, masivamente paralelo y por tanto, de difícil descomposición, lo que requiere un tipo de modelización en términos de teoría de sistemas dinámicos, el cual, en definitiva, está más próximo al enfoque nomológico deductivo que al mecanicista.

Conclusiones

Lo peculiar es pues la imbricación de procesos que requieren localmente una explicación mecanística, con otros, más globales, concurrentes y holísticos, cuyo comportamiento suele generar "patrones emergentes" y fenómenos de "auto-organización".

El estudio de la complejidad biológica requiere, en consecuencia, un replanteamiento profundo de los enfoques y metodologías desarrolladas hasta hoy, buscando una integración de la metodología mecanicista, basada en el análisis de partes funcionales y la reconstrucción de comportamientos, y la modelización predictiva, basada en el descubrimiento de formas de regularidad (leyes o principios generales). Con ello es de esperar no sólo un mayor rigor científico en el estudio de los sistemas biológicos sino también una universalización de los principios que los rigen: es decir, una confluencia entre la biología de la vida que conocemos, y las nacientes CC de la Biología Sintética y la Astrobiología. Es decir, una Biología Universal.

Agradecimientos

El autor agradece la ayuda de los Proyectos FFI2011-25665 y FFU2009-12895-CO2-02, del Ministerio de Ciencia e Innovación y IT 505-10 del Gobierno Vasco.

REFERENCIAS

Bechtel, W. and Abrahamsen, A. (in press). *Thinking dynamically about biological mechanisms: Networks of coupled oscillators*. Foundations of Science.

Mattick, J.S. 2004, *The hidden genetic program of complex organisms,* Scientific American, (Oct.), 61–67

Moreno, A. Ruiz Mirazo, K, & Barandiaran, X. 2011, "The impact of the paradigm of complexity on the foundational frameworks of biology and cognitive science", In *Handbook of the Philosophy of Science,* Volume 10: Philosophy of Complex Systems. Volume editor: Cliff Hooker. General editors: Dov M. Gabbay, Paul Thagard and John Woods. Elsevier pp 311-333.

DESDE LOS SISTEMAS A LOS MECANISMOS DE PROCESAMIENTO DE INFORMACIÓN

PASCUAL F. MARTÍNEZ-FREIRE
Universidad de Málaga
martinez.freire@gmail.com

1. Ciencias cognitivas y sistemas de procesamiento de información

Tanto desde el punto de vista filosófico como desde el punto de vista de los logros científicos, la principal contribución de las ciencias cognitivas (básicamente psicología e inteligencia artificial), consideradas globalmente, ha sido la ampliación y generalización del concepto de proceso mental. En efecto, dentro del campo de investigación de las ciencias cognitivas se estiman como procesos mentales no sólo los relativos a humanos y a animales en general, sino también los relativos a computadores y robots (adecuadamente programados).

Tal contribución ha tenido como base teórica la idea de sistema de procesamiento de información (SPI), elaborada por Allen Newell (1927-1992) y Herbert Simon (1916-2001) en su obra conjunta *Human Problem Solving* (1972). En efecto, en esta obra sus autores dicen que un SPI es un sistema que consta de una memoria que contiene estructuras de símbolos, de un procesador, de efectores y de receptores (Newell & Simon 1972, 20-21). Podemos y creo que debemos entender por "símbolo" cualquier ítem o elemento de carácter representacional, pero lo importante para los autores es que los símbolos constituyen estructuras de símbolos conectados mediante un conjunto de relaciones. A su vez, la memoria es el componente de un SPI capaz de almacenar y retener estructuras de símbolos. Por otra parte, un proceso de información es un proceso que tiene estructuras de símbolos en (algunas de) sus entradas y salidas, mientras que el procesador es el componente de un SPI que consta de un conjunto (fijo) de procesos de información elementales, de una memoria de corto plazo (MCP) que mantiene las estructuras de símbolos de entradas y salidas, y de un intérprete que determina las secuencias de procesos de información a ejecutar por el SPI. Me parece que conviene destacar que una estructura de símbolos designa, se refiere a, o señala a un objeto en el caso de que existan procesos de información que, por un lado, admiten la estructura de símbolos como entrada y, por otro lado, o afectan al objeto o producen como salida estructuras de

símbolos que dependen del objeto. La indicación de la designación o referencia en un SPI por parte de Newell y Simon resulta central, ya que podría criticarse la noción de SPI, dada su caracterización formal y abstracta, señalando que sólo tiene aspectos formales y sintácticos, pero en realidad queda claro en la postura de sus autores que las estructuras de símbolos pueden designar objetos del ambiente y, de hecho, la figura 2.1 del texto que comento incluye el ambiente. Asimismo una estructura de símbolos es un programa si tanto el objeto que designa es un proceso de información como, dado el programa, el intérprete (del procesador) puede ejecutar el proceso designado. Finalmente un símbolo es primitivo si su designación o su creación está fijada por procesos de información elementales o bien por el ambiente externo del SPI. (Obsérvese de nuevo la indicación del ambiente).

Debe quedar claro que un SPI constituye una descripción abstracta de un sujeto cognitivo, incluyendo tanto los conocedores humanos como los computadores. Por ello, en el epílogo de su obra Newell y Simon declaran que el punto de vista básico que ha dominado su obra ha sido que el computador programado y el solucionador de problemas humano son ambos especies pertenecientes al género SPI, aunque por supuesto difieren en organización de memoria, en procesos elementales y en organización de programas (Newell & Simon 1972, 870). Cabe añadir que, atendiendo al desarrollo del estudio de la mente animal, tal como aparece de modo ejemplar en la obra de Donald Griffin (1915-2003) titulada *Animal Minds* (1992), podemos distinguir tres especies de SPI, que son el conocedor humano, el conocedor animal (al menos los vertebrados) y el computador adecuadamente programado.

2. Funcionalismo y cognición corporizada

Si bien las ciencias cognitivas (psicología e inteligencia artificial) ya están unidas por la noción de SPI, esta unión, a mi entender, queda consolidada por la hipótesis filosófica del funcionalismo respecto de la naturaleza de los procesos mentales. El funcionalismo en filosofía de la mente debe mucho tanto a Hilary Putnam (Putnam 1967) como a Jerry Fodor (Fodor 1968). El gran mérito de Putnam estriba en haber defendido, frente al materialismo o teoría de la identidad entre procesos mentales y procesos neuronales, que los estados mentales son estados funcionales de un organismo tomado en su totalidad. A su vez, el principal mérito de Fodor consiste en distinguir claramente entre función y su soporte. Tal como he señalado en "La interdisciplina en filosofía de la mente" (2011a), por funcionalismo debe entenderse la doctrina de que los procesos mentales no son conducta externa públicamente observable (como defendía el conductismo), sino que son funciones mediadoras entre entradas sensoriales y salidas motoras, que se producen en el interior del sistema y que son causa de la conducta externa. Los procesos mentales son funciones en tres sentidos. En

primer lugar, son funciones mediadoras entre estímulos del ambiente y respuestas del sistema. En segundo lugar, son funciones dependientes tanto de los estímulos informativos como de la información presente en el agente y su disposición en éste. Y en tercer lugar, son funciones caracterizadoras de un papel o rol causal.

Ahora bien, el funcionalismo cognitivo significa la distinción entre función psicológica y su soporte. Al respecto David Marr (1945-1980) distingue tres niveles en los que un mecanismo (*device*) de procesamiento de información debe ser entendido. En un extremo, el nivel superior, está la teoría computacional abstracta del mecanismo, donde se especifica el objetivo de la computación; en el centro, está la elección de representación para entradas y salidas y para el algoritmo que transforma unas en otras; y al otro extremo están los detalles de cómo el algoritmo y la representación se realizan físicamente (Marr 1982, 24-25). Dicho en términos simples, además de especificar las tareas del mecanismo, cabe distinguir entre la descripción funcional del mecanismo y la descripción del soporte del mecanismo.

Hilary Putnam imagina que hay un científico diabólico que se dedica a quitar los cerebros de los cuerpos humanos y colocarlos en una cubeta con nutrientes que mantienen los cerebros vivos; además las terminaciones nerviosas de estos cerebros se conectan con un supercomputador que produce en los cerebros la ilusión de que todo sigue igual; en un paso más, Putnam imagina que todos los seres humanos somos cerebros en una cubeta con nutrientes y conectados a un supercomputador como el descrito, y entiende que esta suposición puede ayudar a comprender la relación mente/mundo (Putnam 1981, 5-6). Al margen de que esta ficción sirva o no para algo, lo cierto es que Larry Shapiro parece tomársela en serio, al señalar que en la ciencia cognitiva dominante se defiende el *encubatamiento* (*envatment*), es decir, se sostiene que la mente, como el proverbial cerebro en una cubeta, es una entidad distinta del cuerpo y tiene como su función el procesamiento de señales desde el cuerpo, y la computación y transmisión de señales de vuelta al cuerpo (Shapiro 2004, 169).

Sin embargo, resulta completamente aventurado e irreal atribuir a la posición funcionalista o tradicional de las ciencias cognitivas la concepción de que la mente *es* una entidad separada del cuerpo. En realidad, lo que sostiene el funcionalismo cognitivo es la distinción básica entre función mental y soporte de tal función. Por ejemplo, la función puede ser tipificar la información obteniendo un concepto, y el soporte puede ser un cerebro humano, un cerebro animal o un "cerebro" electrónico.

3. Mecanismos de procesamiento de información

En todo caso, y a fin de evitar cualquier acusación de "descorporización" o de "nuevo cartesianismo" (en el sentido de defender que los procesos mentales son espirituales), creo que deben sostenerse tres tesis, que no son incompatibles con el funcionalismo cognitivo.

En primer lugar, al distinguir entre función mental y su soporte no se defiende que pueda *darse* una función mental sin soporte alguno, sino simplemente que cabe *caracterizar* un proceso mental prescindiendo *en principio* de su soporte. Algo parecido ocurre cuando caracterizamos la respiración como la función de captación de oxígeno, que tiene como posibles soportes tanto pulmones como branquias o tráqueas. Y de la misma manera que a nadie se le ocurre decir que, *en los seres humanos*, la respiración no tiene nada que ver con los pulmones, tampoco a nadie se le ocurre decir que, *en los seres humanos*, los procesos mentales no tienen nada que ver con el cerebro.

En segundo lugar, es necesario incluir suficientes referencias al soporte cuando se describe el mecanismo que produce una función concreta. Como vimos antes, David Marr utilizaba el término "mecanismo" (*device*), pero también escribió: "Así pues el punto crítico es que comprender computadores es diferente de comprender computaciones. Para comprender un computador, uno tiene que estudiar ese computador. Para comprender una tarea de procesamiento de información, uno tiene que estudiar esa tarea de procesamiento de información. Para comprender plenamente una máquina particular realizando una tarea particular de procesamiento de información, uno tiene que hacer ambas cosas. Ninguna de ellas sola bastará" (Marr 1982, 5). Por otra parte, cuando describimos funciones psicológicas determinadas de un animal (comprendido el humano) concreto, son necesarias referencias suficientes a los mecanismos neuronales que las producen.

En tercer lugar, conviene distinguir, tal como hace William Bechtel, entre las operaciones de las partes de un mecanismo y la función del mecanismo total (Bechtel 2008, 13). La expresión "mecanismo de procesamiento de información" (MPI) es explícitamente empleada por Bechtel, quien, refiriéndose a René Descartes (1596-1650), alaba con razón por un lado su instauración de las explicaciones mecanicistas, mientras que condena por otro lado su dualismo entre procesos mentales y procesos cerebrales. Para Bechtel, un mecanismo consta de partes que ejecutan operaciones de una manera organizada y es la cooperación entre las operaciones de las partes la que da razón de la función del todo. Por otro lado me parece importante destacar que, para Bechtel, es central a la idea de un mecanismo de procesamiento de información que los estados dentro del mecanismo sirven para llevar información acerca de objetos o eventos externos al mecanismo; el término "representar" (*represent*) caracteriza la relación entre estados dentro del sistema [*sic*] de procesamiento de información, que son los vehículos de la representación, y los objetos o eventos

externos, que son los contenidos de la representación (Bechtel 2008, 159).

4. El nuevo proyecto de mecanismos de procesamiento de información en ciencias cognitivas

Así entendidas las cosas, la vieja idea del sistema de procesamiento de información (SPI) puede ser refinada en la nueva idea del mecanismo de procesamiento de información (MPI). A mi entender hay aquí un cambio de acento que es de la mayor importancia. Ciertamente abstraer de algo no es rechazar la existencia de ese algo, pero cuando asumimos descripciones abstractas siempre corremos el peligro de menospreciar los elementos concretos, o al menos siempre corremos el peligro de ser acusados de tal menosprecio. Esto ha ocurrido con la noción de SPI, que ha sido atacada desde las instancias de la cognición situada o de la cognición corporizada. La cognición situada, cuyo principal representante es probablemente la antropóloga Lucy Suchman, insiste en que todas las acciones humanas son acciones realizadas en el contexto de circunstancias concretas y particulares, es decir, son acciones situadas. A su vez, la cognición corporizada, cuyo principal exponente es quizás el citado Larry Shapiro, insiste en que el cuerpo contribuye e influye en las actividades cognitivas en una medida que la ciencia cognitiva no ha apreciado adecuadamente. En mi opinión, en la base de ambas instancias se encuentran las ideas que presentó Hubert Dreyfus en sus célebres críticas de la inteligencia artificial. Dreyfus publicó en 1972 su libro *What Computers Can´t Do*, el cual tuvo una edición revisada en 1979 (con una nueva introducción), y tuvo una nueva versión en 1992 (con otra nueva introducción) con el nuevo título de *What Computers Still Can´t Do*. Para este filósofo, los seres humanos tienen (al menos) dos características que no pueden ser imitadas por los computadores, a pesar de que subyacen a toda conducta inteligente, a saber, el papel de la situación en que nos encontramos y el papel del cuerpo en la organización y unificación de nuestra experiencia de los objetos. Es más el propio Dreyfus emplea la expresión "agente corporizado" (*embodied agent*) para referirse a los agentes humanos (Dreyfus 1992, 255).

Pero estas críticas no son de recibo. Por una parte, tal como se indicó antes, un SPI es un agente cognitivo situado en un ambiente y sus receptores reciben información del ambiente (no de un supercomputador, como en la ficción científica citada de Putnam). Y por otra parte, tal como he defendido en mi estudio "La Unidad de las Ciencias Cognitivas: Crítica de la ´Cognición Incorporizada`" (2011b), hablar de "cognición corporizada" (o "incorporizada") es o bien un pleonasmo o bien una contradicción. Si nos referimos a la cognición animal en general y humana en particular, resulta obvio que los animales y los humanos son realidades corporales, con lo que añadir "corporizada" a "cognición" constituye un caso de pleonasmo (como añadir "abajo" a "bajar"). Si nos referimos a la cognición de los computadores, también resulta obvio que los computadores no tienen un cuerpo biológico, con lo que añadir "corporizada" a "cognición"

constituye un caso de contradicción.

De todas formas al hablar de mecanismos de procesamiento de información (MPI), en vez de sistemas de procesamiento de información (SPI), estamos introduciendo un nuevo proyecto que busca resultados explícitamente más concretos y un estudio detallado de los soportes de las funciones mentales.

Esta estrategia mecanicista está siendo justificada y animada por filósofos de la ciencia contemporáneos en relación con los estudios psicológicos. Tal es el caso de Cory Wright y William Bechtel en su trabajo " Mechanisms and Psychological Explanation" (2007). Para estos autores, actualmente es práctica explicativa común en psicología la descomposición de un sistema compuesto, organizado jerárquicamente, en sus partes y operaciones componentes, y luego la construcción de modelos que incitan la comprensión científica de cómo podrían estar organizadas para componer la actividad del mecanismo. En vez de subsumirse bajo conjuntos de leyes, tales modelos figuran en narraciones acerca de cómo un mecanismo puede ser directamente responsable de un fenómeno (Wright & Bechtel 2007, 72).

A su vez, esta estrategia mecanicista resulta mucho más normal en inteligencia artificial. Desde siempre en la construcción de computadores, sus partes son diseñadas para realizar operaciones que la unidad de control organiza a fin de obtener un resultado global, que es la función del computador. La roboticista Maja Mataric señala que la arquitectura de un computador en general, como la de un robot en particular, significa el conjunto de principios para diseñar computadores a partir de una colección de elementos constituyentes (*building blocks*) bien entendidos. Estos elementos constituyentes o partes son instrumentos (*tools*) a nuestra disposición para hacer que la tarea de diseño del control del computador o robot sea más fácil (Mataric 2007, 136). En suma, la estrategia mecanicista en inteligencia artificial intenta comprender la función del mecanismo total considerando las operaciones de las partes, que justamente han sido diseñadas como instrumentos del mecanismo total.

REFERENCIAS

Bechtel, W. 2008, *Mental Mechanisms. Philosophical Perspectives on Cognitive Neuroscience*, Nueva York, Taylor & Francis Group.

Dreyfus, H. 1992, *What Computers Still Can´t Do. A Critique of Artificial Reason*, Cambridge (MA), The MIT Press.

Fodor, J. A. 1968, *Psychological Explanation. An Introduction to the Philosophy of Psychology*, Nueva York, Random House.

Griffin, D. R. 1992, *Animal Minds*, Chicago, The University of Chicago Press.

Marr, D. 1982, *Vision. A Computational Investigation into the Human Representation and Processing of Visual Information*, Nueva York, W. H. Freeman and Company.

Martínez-Freire, P. F. 2011a, "La interdisciplina en filosofía de la mente", *Ludus Vitalis*, 35, 227-

230.

Martínez-Freire, P. F. 2011b, "La Unidad de las Ciencias Cognitivas: Crítica de la ´Cognición Incorporizada`", en Pombo, O. (ed.), *Lógica Universal e Unidade da Ciência*, Lisboa, Universidade de Lisboa, 43-61.

Mataric, M. J. 2007, *The Robotics Primer*, Cambridge (MA), The MIT Press.

Newell, A. y Simon, H. 1972, *Human Problem Solving*, Englewood Cliffs (NJ), Prentice-Hall.

Putnam, H. 1967, "Psychological Predicates", en Capitan, W. H. y Merrill, D. D. (eds.), *Art, Mind and Religion*, Pittsburgh, University of Pittsburgh Press, 37-48.

Putnam, H. 1981, *Reason, Truth and History*, Nueva York, Cambridge University Press.

Shapiro, L. 2004, *The Mind Incarnate*, Cambridge (MA), The MIT Press.

Wright, C. y Bechtel, W. 2007, "Mechanisms and Psychological Explanation", en Thagard, P. (ed.), *Philosophy of Psychology and Cognitive Science*, Amsterdam, North-Holland, 31-79.

EXPLICACIÓN Y COMPRENSIÓN EN BIOLOGÍA: EL USO DE MODELOS

Antonio Diéguez
Universidad de Málaga
dieguez@uma.es

Es ampliamente aceptado que una de las funciones principales de los modelos científicos es proporcionar una explicación de los fenómenos que representan. Se dice, por ejemplo, que el modelo operón de la lactosa explica cómo funciona la regulación genética en algunas circunstancias; que el modelo Lotka-Volterra de competencia interespecífica explica los resultados posibles de la competición entre dos especies por el mismo nicho ecológico; que el modelo de Fisher de la ratio entre sexos explica por qué, en condiciones habituales, la proporción entre machos y hembras es de 1:1; que el modelo de doble hélice del ADN explica los patrones de difracción de rayos X detectados por Rosalind Frankin. Sin embargo, a pesar de algunos avances notables en las últimas décadas, la cuestión de cómo explican los modelos sigue siendo muy controvertida.

La situación se torna particularmente apremiante en el caso de la biología, ya que, si bien el recurso habitual para la formación de explicaciones en física y en química, que son las leyes científicas, es (prácticamente) inexistente en biología, en cambio, la biología emplea constantemente modelos como herramientas explicativas. Se ha dicho incluso que "la explicación en biología es obtenida siempre a través de la intervención directa de modelos de los fenómenos a explicar" (Leonelli 2009, p. 192). Es, pues, razonable preguntarse si lo que es válido según los diversos modelos de explicación científica, y muy particularmente el modelo de explicación por cobertura legal, es aplicable de algún modo a la explicación mediante modelos.

En esta comunicación se discutirán algunos problemas en relación con la posibilidad de explicar mediante modelos en biología. Se argumentará que los modelos tradicionales de explicación científica no encajan bien con el modo en que se realiza la explicación mediante modelos. Los modelos pueden ofrecer explicaciones de formas muy diversas, entre otras razones porque hay muchos tipos de modelos, de modo que, muy posiblemente no sea factible reducir a un patrón único la explicación mediante modelos. No obstante, si hay un rasgo común a estos

diferentes modos de proporcionar explicación, y que, por tanto, merece un análisis detenido, éste es el logro de una mejor comprensión de los fenómenos (cf. De Regt *et al*. 2009).

Se argumentará que la comprensión no es una mera consecuencia o un subproducto de la explicación, sino que, por el contrario, proporcionar comprensión (de diferentes formas, dependiendo del contexto) es el fin mismo de cualquier explicación mediante modelos. Esta noción de comprensión no es, como a veces se objeta, irremediablemente subjetiva. Pueden proporcionarse algunos criterios justificables para establecer cuándo un sentimiento subjetivo de comprensión se corresponde con una comprensión genuina. Este problema no es en el fondo más irresoluble que el problema clásico de distinguir entre conocimiento y creencia. Sin embargo, estos criterios no pueden ser los mismos para todo tipo de modelos. En particular, es necesario distinguir entre "modelos contrastantes" y "modelos representacionales".

La diversidad de modelos de explicación científica que han sido propuestos en las últimas décadas –en algunos aspectos muy distantes unos de otros– no debe ocultar que hay algunos puntos acerca de la explicación científica en los que se da un amplio consenso. Son muchos los autores que coinciden en señalar que la explicación científica de un fenómeno consiste en proporcionar una respuesta a una pregunta por-qué y esa respuesta debe incluir por lo general una identificación de la causa del fenómeno o, al menos, la incorporación del fenómeno a una estructura causal. Por otro lado, aunque hay varios disidentes, se sostiene habitualmente que una explicación científica debe incluir como elemento indispensable el recurso a una ley científica o, al menos, de una generalización que permanezca invariable bajo algunas intervenciones (cf. Wookward 2001). Si trasladamos esto al modo en que un modelo científico puede proporcionar una explicación de un fenómeno, tendríamos entonces que concluir que sólo los modelos que contestan a preguntas por-qué y que contienen leyes o que son ellos mismos leyes de cierto tipo podrían ser considerados como explicativos. Sin embargo, si asumimos, con la mayoría de los filósofos de la biología, la inexistencia (o escasez) de leyes en la biología, entonces deberíamos aceptar asimismo que esta ciencia apenas tiene capacidad para proporcionar explicaciones. Y, no obstante, lo que observamos es que muchos modelos biológicos que no contienen leyes (biológicas) son considerados habitualmente como explicativos. Por ejemplo, el modelo de bicapa lipídica de la membrana celular, el modelo quimiosmótico de Peter Mitchell de la fosforilación oxidativa, o el modelo de Rice y Salt de la especiación simpátrica.[1]

[1] Podría argüirse que el modelo de van Frasssen –que no exige que la explicación sea un argumento ni que haga uso de una ley científica– podría ser adaptado para describir cómo explican los modelos biológicos. Sin embargo, creo que este no es el mejor camino a tomar. El modelo pragmático de explicación de van Fraassen sigue siendo demasiado restrictivo. En primer lugar, no todo modelo es una respuesta a una pregunta por-qué. En segundo lugar, los modelos no son siempre capaces de establecer una clase de contraste en cualquier asunto. Y, finalmente, los criterios que van

Una dificultad importante en el análisis de esta cuestión la ofrece la polisemia del término 'modelo'. En biología, en particular, 'modelo' puede designar cosas muy diferentes: organismos concretos (como *Drosophila melanogaster*) que, por diversos motivos, son ampliamente usados en la investigación; objetos materiales que representan de forma simplificada a otras entidades (como los modelos moleculares hechos con plástico y metal); soluciones paradigmáticas a problemas empíricos (como el modelo de operón de la lactosa); interpretaciones teóricas e idealizadas de la estructura y mecanismos funcionales de algunas entidades o procesos biológicos (como el modelo de bicapa lipídica, el modelo llave-cerradura de la acción enzimática, el modelo de McArthur y Wilson de la biogeografía insular, el modelo de Mitchell de fosforilación oxidativa); conjunto de ecuaciones que describen algunos aspectos del comportamiento de sistemas biológicos complejos (como el modelo Lotka-Volterra de competencia interespecífica, el modelo cinemático Michaelis-Menten para la determinación de la velocidad de la acción enzimática, o el modelo metapoblacional de Levins); programas de simulación en ordenador (como los programas de Vida Artificial); etc. Demetris Portides (2008, p. 385) ha señalado que la mayor parte de esto sentidos de 'modelo' están ligados a las ideas de representación e idealización. Pero otros autores niegan incluso este mínimo común denominador. ¿Puede encontrarse, entonces, algún elemento común a las explicaciones proporcionadas por estos diversos tipos de modelos?

En general, puede afirmarse que, tanto en la ciencia como en la vida cotidiana, se logra una explicación de un hecho cuando este hecho es incorporado a un cuerpo previo de conocimientos mediante conexiones que muestren que el hecho es causalmente posible dado ese cuerpo de conocimientos. La concepción heredada de la explicación científica capturaba esta intuición mediante la idea de que pedir una explicación científica de un fenómeno era equivalente a preguntar *por qué* sucedió el fenómeno. Sin embargo, es fácil ver que muchos modelos en la ciencia no pueden interpretarse como buscando respuestas a preguntas por-qué (*why-questions*), sino más bien a preguntas cómo-es-posiblemente (*how-possibly-questions*), cómo-es-de-hecho (*how-actually*) y cómo-es-plausible-que-sea (*how-plausibly*) (cf. Salmon 1989, pp. 136-138, Machamer, Darden & Craver 2000, and Craver 2006).

Por ello, propongo entender la explicación científica en un sentido amplio como una respuesta a una pregunta (no necesariamente a una pregunta por-qué) suscitada en un contexto científico que proporciona una *mejor comprensión* de algún fenómeno o conjunto de fenómenos. Esta definición amplia permite incluir fácilmente a los modelos como herramientas explicativas. Los modelos explican porque pueden proporcionar este tipo de respuestas. Si hay algo en común en

Fraassen propone para evaluar las respuestas dadas en una explicación establecen condiciones muy estrictas –tal como favorecer una clase de contraste en lugar de otra– que muchos modelos explicativos no pueden satisfacer.

el modo en que explican los diferentes tipos de modelos es esto: la explicación aumenta nuestra comprensión del sistema representado por el modelo. Esta conexión entre explicación y comprensión no es nada nuevo. Frente a la tradición historiográfica antipositivista decimonónica y a la filosofía hermenéutica, que ha tendido a contraponer la explicación científica a la comprensión de acciones y significados, son muchos los autores que han conectado ambos conceptos (cf. Scriven 1962, Friedman 1974, Kitcher 1981, Achinstein 1983, Kim 1994, Salmon 1998, Faye 1999).

Si nos centramos en los modelos en lugar de en las teorías, parece haber diversos modos –y no uno sólo– en los que éstos pueden mejorar nuestra comprensión de los fenómenos. Una de las formas más importantes, y posiblemente la más frecuente, en las que puede hacerlo es describiendo, de una forma simplificada, los mecanismos causales que producen el fenómeno. Éste suele ser el caso en los modelos analógicos (cf. Salmon 1998, p. 77, Machamer, Darden & Craver 2000, Harré 2004, p. 76, Bechtel & Abrahamsen 2005, Craver 2006, Darden 2008, Bechtel 2011). Pero no debe pensarse que esta es la única forma en la que lo hacen. Otra forma de hacerlo es poniendo de relieve por medio de ecuaciones matemáticas las relaciones estructurales entre los rasgos fundamentales del fenómeno representado. Otra, que ha sido defendida por Michael Friedman y por Philip Kitcher, es mediante la unificación de fenómenos constitutivamente diferentes. Otra, como en el modelo de Fisher de proporción de sexos, es mostrando cómo diversos escenarios causales con diferentes condiciones iniciales conducen sin embargo al mismo resultado. En este asunto, como en tantos en filosofía de la biología, el enfoque pluralista se torna necesario.

A pesar de la importancia de este asunto, el análisis preciso de la noción de comprensión científica en relación con los fenómenos naturales está aún en sus comienzos. Y si quiere ser de alguna utilizad para la clarificación de la noción de explicación, ha de enfrentarse a algunas dificultades. Entre ellas no es la menor la que hace años ya señalara Hempel (1965, pp. 425-6), a saber: que la comprensión es un asunto meramente subjetivo, una noción atribuible sólo a estados subjetivos de individuos concretos, y por tanto un concepto relativo y no generalizable. Newton-Smith ha ilustrado el problema con un bello ejemplo:

> Las definiciones de los diccionarios suelen explicar la noción de explicación en términos de comprensión: una explicación es algo que proporciona una comprensión o torna algo inteligible. Quizás esta es la noción unificadora. Los tipos diferentes de explicación son tipos de explicación en virtud de su capacidad para proporcionar comprensión. Sin embargo, aunque ciertamente una explicación ha de ser capaz de proporcionarle a una persona adecuadamente educada un sentido psicológico de comprensión, no es probable que este sea un camino fructífero, puesto que no hay virtualmente ningún límite a lo que ha sido aceptado como capaz de proporcionar una comprensión. Hubo un tiempo en que muchos pensaban que el hecho de que hubiera siete virtudes y siete orificios en la cabeza de los seres humanos les proporcionaba una comprensión de por qué (supuestamente) había sólo

> siete planetas. Necesitamos distinguir entre comprensión real y espuria. Y para esto necesitamos una teoría filosófica de la explicación que nos diga qué es lo que distingue a una buena explicación. (Newton-Smith 2001, p. 131).

El problema, por tanto, consiste en averiguar si hay algún criterio para distinguir entre la mera *sensación de comprensión* y la comprensión genuina. Éste es, de hecho, un viejo problema: fue la crítica principal que recibió el método de *Verstehen* en el debate clásico en metodología de las ciencias sociales.

Ciertamente, es imposible proporcionar un conjunto exhaustivo y permanente de criterios que permita señalar cuándo estamos ante una comprensión genuina, puesto que estos criterios poseen elementos contextuales que han cambiado a lo largo de la historia de la ciencia (de Regt & Dieks 2005). No obstante, creo que sí pueden ofrecerse algunas indicaciones útiles que permitan decidir de forma razonable si un modelo científico nos proporciona o no una comprensión genuina. Para ello hemos de comenzar por distinguir entre dos tipos diferentes de modelos a los que denominaré '*modelos contrastantes*' y '*modelos representativos*'.

Entiendo por modelos contrastantes aquéllos construidos intencionalmente como recursos explicativos extremadamente no-realistas, como es el caso de los modelos falsos en biología (e.g. poblaciones con más de dos sexos, poblaciones con herencia mezclada en lugar de mendeliana). Son modelos que no tratan de representar ningún sistema real y que no se hacen más precisos mediante un proceso de des-idealización. El sistema representado es completamente ficticio, aunque es propuesto para entender algunos aspectos de un sistema-diana real. En este tipo de modelos, la carga explicativa reside justamente en las disimilitudes que el modelo presenta con el sistema-diana.

Para este tipo de modelos, propongo un criterio cualitativo simple de comprensión genuina:

> Un modelo contrastante nos proporciona una comprensión genuina del comportamiento de un sistema real si el contraste entre las consecuencias lógicamente derivables del modelo permite revelar cómo algunas características interesantes del comportamiento del sistema real dependen de la presencia o ausencia de ciertas circunstancias que están ausentes o presentes respectivamente en el modelo.

Por modelos representativos entiendo aquellos que encajan de algún modo con el sistema-diana, o, en otros términos, aquellos que son construidos con el propósito de representar algunos aspectos de los fenómenos presentes en un sistema real.

Para este tipo de modelos, dada su variedad, es más difícil seleccionar criterios que sirvan como indicadores de comprensión genuina. No obstante, creo que los siguientes, puramente negativos, pueden ser una base mínima para tales criterios, permitiendo excluir algunos modelos:

Un modelo representacional *no* proporciona una comprensión genuina del sistema-diana si:

(1) Las analogías entre el modelo y el sistema-diana son débiles, *ad hoc*, o científicamente infundadas.
(2) Formula abstracciones sobresimplificadas que excluyen factores funcionales relevantes, esto es, factores que son necesariamente constitutivos del comportamiento del sistema modelado.
(3) Realiza idealizaciones extremadamente irreales o inútiles, esto es, idealizaciones que están tan alejadas de las condiciones con las que ha de enfrentarse el sistema modelado que, incluso tras correcciones progresivas, no ayudan a ver cómo este sistema varía en las circunstancias usuales bajo algunas manipulaciones.
(4) Postula una ontología fantástica o pseudocientífica.
(5) Los mecanismos postulados carecen de analogías con los mecanismos que funcionan en el sistema real.
(6) Sus predicciones adicionales fracasan sistemáticamente.

Usando la terminología de Weisberg (2007), podríamos considerar que éstos son criterios mínimos negativos de "fidelidad representacional".

Conclusiones

La comprensión no es un mero subproducto de una buena explicación, sino que, en lo que a la explicación mediante modelos concierne, puede considerarse que es el objetivo mismo de la explicación. Una buena explicación de un fenómeno es aquella capaz de proporcionar a la comunidad científica relevante una mejor comprensión del fenómeno. La noción de comprensión no es irremediablemente subjetiva. Pueden proporcionarse algunos criterios contextuales, pero objetivamente justificables, para distinguir entre la mera sensación de comprensión y la comprensión genuina.

REFERENCIAS

ACHINSTEIN, P. 1983: *The Nature of Explanation*. New York: Oxford University Press.
BECHTEL, W. 2011: "Mechanism and Biological Explanation", *Philosophy of Science*. Vol. 78, No. 4, pp. 533-557.
BECHTEL, W. & ABRAHAMSEN, A. 2005: "Explanation: A Mechanist Alternative", *Stud. Hist. Phil. Biol. & Biomed. Sci*., 36, pp. 421-441.
CRAVER, C. F. 2006: "When Mechanistic Models Explain", *Synthese*, 153, pp. 355-376.

DARDEN, L. 2008: "Thinking Again about Biological Mechanisms", *Philosophy of Science*, 75, pp. 958-969.

DE REGT, H. W. & DIEKS, D. 2005: "A Contextual Approach to Scientific Understanding", *Synthese*, 144, pp. 137-170.

DE REGT, H. W., LEONELLI, S. & EIGNER, K. 2009: "Focusing on Scientific Understanding", en H. W. de Regt, S. Leonelli & K. Eigner (Eds.) *Scientific Understanding: Philosophical Perspectives* (pp. 1-17). Pittsburgh: University of Pittsburgh Press.

FAYE, J. 1999: "Explanation Explained", *Synthese*, 120, pp. 61-75.

FRIEDMAN, M. 1974: "Explanation and Scientific Understanding", *The Journal of Philosophy*, 71, pp. 5-19.

HARRÉ, J. 2004: *Modeling: Gateway to the Unknown*. (Edited by D. Rothbart), Amsterdam; Boston: Elsevier.

HEMPEL, C. G. 1965: *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

KIM, J. 1994: "Explanatory Knowledge and Metaphysical Dependence", *Philosophical Issues*, 5, pp. 51-69.

KITCHER, Ph. 1981: "Explanatory Unification", *Philosophy of Science*, 48, pp. 507-531.

LEONELLI, S. 2009: "Understanding in Biology: The Impure Nature of Biological Knowledge", en H. W. de Regt, S. Leonelli & K. Eigner (Eds.) *Scientific Understanding: Philosophical Perspectives* (pp. 189-209). Pittsburgh: University of Pittsburgh Press.

MACHAMER, P. K., DARDEN, L., CRAVER, C. F. 2000: "Thinking about Mechanisms", *Philosophy of Science*, 57, pp. 1-25.

NEWTON-SMITH, W. H. 2001: "Explanation", en W. H. Newton-Smith (ed.) *A Companion to the Philosophy of Science*. Malden, MA: Blackwell, pp. 127-133.

SALMON, W. 1989: "Four Decades of Scientific Explanation", en Ph. Kitcher and W. Salmon (eds.) *Scientific Explanation. Minnesota Studies in the Philosophy of Science, vol. XIII*, Minneapolis: University of Minnesota Press, pp. 3-219.

SCRIVEN, M. 1962: "Explanations, predictions, and laws", en H. Feigl and G. Maxwell (eds.) *Scientific explanation, space, and time*. Minneapolis: University of Minnesota Press, pp. 170-230.

WEISBERG, M. 2007: Who is a Modeler?, *British Journal for the Philosophy of Science*, 58, 207-233.

WOODWARD, J. 2001: "Law and Explanation in Biology: Invariance is the Kind of Stability That Matters", *Philosophy of Science*, 68, pp. 1-20.

CONSIDERACIONES CRÍTICAS ACERCA DEL CONCEPTO DE "NIVELES DE MECANISMOS"

JON UMEREZ
Universidad del País Vasco
jon.umerez@ehu.es

ABSTRACT: This work is an attempt to clarify the concept of level in order to make it viable to use it coherently within different conceptual frames and, especially, to allow discussion and commensurability among diverse and even alternative views related, in particular, to biological systems. Instead of proposing *a* definition I assume an intuitive notion in the sense that levels are those diverse states of association of elements that constitute a system by virtue of a given **relation**. Which that relation will be is precisely what demands a detailed epistemological reconstruction. The features considered are *composition*, *integration*, *emergence*, *control* and *organization*.
PALABRAS CLAVE:
Causación descendente, constricción, mecanicismo, niveles de organización, relación internivélica

Introducción

Las consideraciones críticas que se presentan aquí se fundan en un trabajo de clarificación epistemológica del concepto de nivel con el objetivo de hacerlo viable para su uso coherente en marcos conceptuales diferentes y que permita la discusión y la conmensurabilidad entre perspectivas diversas e incluso alternativas referidas a los sistemas biológicos. No obstante su generalidad, el análisis se orienta a contribuir a una explicación de la organización biológica con la elaboración del concepto de *nivel de organización*. En este sentido, no se trata de obtener una definición o una taxonomía definitivas sino de tomar una noción intuitiva como punto de partida y dotarla de contenido específico en tanto que interrelación buscando la articulación entre diversos tipos o grados de relación.

Una constatación preliminar es que se trata de un concepto que, normalmente, ha sido estudiado con o a través del concepto *jerarquía*. Es decir, la mayor parte del trabajo se ha hecho acerca de jerarquías, muy poco sobre niveles como tales. Sin embargo, la idea de nivel merece ser objeto de un análisis conceptual separado. De hecho, en este enfoque se plantea que es metodológica y epistemológicamente conveniente proceder desde los niveles a las jerarquías y tomar jerarquía, precisamente, como "relación inter-nivel" en vez de tratar los niveles como algo dado.

1. Noción intuitiva.

Como punto de partida, los niveles son aquellos diversos estados de asociación de elementos que están conectados con otros estados de asociación de elementos en virtud de una determinada relación y dentro de un sistema particular. La única restricción que implica este concepto intuitivo es que la constitución de los sistemas haya de estar basada en algún tipo de **relación específica** entre asociaciones de elementos. Cuál será esa relación en cada caso particular o acepción del concepto de nivel es precisamente lo que demanda una reconstrucción epistemológica detallada. Tomado como un punto de partida en vez de cómo una definición o una taxonomía. Con un concepto de este tipo una definición corre el riesgo de acotar un sentido restringido y confinado, una taxonomía, en cambio, adolece de resultar fallida o no concluyente si no es exhaustiva y, a la vez, resulta casi imposible ser exhaustiva y probablemente inútil si lo es.

2. Ambigüedad y pluralidad de sentidos.

Últimamente disponemos de algunas propuestas muy interesantes para abordar la organización biológica in términos de mecanismos (cuyo origen se puede rastrear en el trabajo de Wimsatt, v.g. 1976, 1994) que han dado lugar al desarrollo de un renovado e idiosincrásico enfoque mecanicista ("mechanistic approach" o "Chicago mechanists" en denominación de Craver). Un elemento importante en este enfoque es su renovado interés por el concepto de nivel. En sus textos a este respecto es habitual encontrar quejas de que el trabajo que se ha hecho acerca del concepto es insuficiente y que la palabra tiene muchas acepciones y sentidos y por lo tanto es difícil de usar y necesita clarificación. Es por ello que, en diversos lugares (p.e., Bechtel 2008, Craver 2007, 2009, Craver & Bechtel 2007) desarrollan y proponen el concepto de *niveles de mecanismos* ("levels of mechanisms") en tanto que noción apropiada para las explicaciones en términos de mecanismos en diversas áreas de las ciencias biológicas y cognitivas.

> The notion of *levels* is ubiquitous in discussions of science. Yet, it is used in such a variety of ways that when it is invoked in discussions of reduction without explication, it is unclear what is to be reduced to what. After surveying some of the most common ways of construing levels and their limitations, I will develop a notion of levels appropriate for mechanistic explanation. (Bechtel 2008, ch. 4, p. 143)

> Explanations in neuroscience typically span multiple levels. The term level, however, is multiply ambiguous. I develop a taxonomy of different kinds of levels, and I show why one must be careful to keep these different kinds distinct. (…) I argue that "levels of mechanisms" captures the central explanatory sense in which explanations in neuroscience (and elsewhere in the special sciences) span multiple levels. (Craver 2007, ch. 5, p. 163)

> The intuition that the world can be ordered into a hierarchy of levels is nearly ubiquitous in cognitive neuroscience and the special sciences generally. However scientists and philosophers use the term ’level’ to describe a wide variety of relations and considerable confusion arises from a failure to keep them distinct. For this reason, it is best not to speak of ‘level’ *simpliciter*, but rather to speak of ‘levels of R’, where R is some explicit specification of a relation between levels or of the relata at different levels. (Craver 2009, p.387)

A pesar de esta última precisión de Craver, la noción que defienden estos autores es estrictamente composicional y la utilizan adicionalmente para distinguir entre relaciones causales y relaciones constitutivas, en función de su naturaleza intra o inter-nivélica, y para derivar algunas conclusiones relevantes acerca del reduccionismo o la causación descendente, entre otras.

Esta queja respecto a la ambigüedad, falta de precisión y consiguiente confusión del término *nivel* parece acompañar sin visos de resolución la reflexión acerca del mismo.

Sin embargo no es totalmente cierto que no se haya trabajado acerca de este asunto: niveles (jerarquías). A finales del siglo XIX y comienzos XX tenemos las propuestas de los teóricos de la Evolución Emergente (Lewes, Lloyd Morgan, Broad, …) y en los años 30 y 40 el desarrollo de la teoría de niveles de integración (en el contexto del organicismo y de los inicios de la ciencia de sistemas) con autores como Needham, Novikoff, Feibleman, Redfield y el propio von Bertalanffy, entre otros. Asimismo, en los años 70 y 80, se encuentra el trabajo sobre teoría de jerarquías (sea en general general o más específicamente en evolución o ecología) con las aportaciones de Pattee, Weiss, Whyte et al, Koestler, Salthe, Collier, Allen & Starr, O’Neill et al., Eldrege, Gould, Grene, etc. Parte de este trabajo ha tenido continuidad hasta el presente (sobre todo en ecología, evolución, biología de sistemas). En resumen, además del trabajo más antiguo (años 30, 40) y el área de ciencia de sistemas, desde finales de los años 60 a mediados de los 80 se ha trabajado mucho en perspectivas organizacionales y evolutivas. Algunos de esos autores (Ahl, y otros ecólogos, Gould, Eldrege y otros biólogos evolutivos, Collier, Salthe …) han continuado defendiendo enfoques jerárquicos y multinivélicos muy específicos en su caracterización del sentido en el que usan “nivel”. Por su parte, Wimsatt (desde los años 70, 1972, 1974, 1976, etc,) inició un enfoque diferente y particular y su vía está siendo seguida hoy en día por otros investigadores. Ahora bien, es también cierto que e término continúa siendo polisémico y que no se ha planteado ninguna definición/teoría general/ admitida/ acordada

3. Enfoque articulado

El objetivo de esta propuesta es complementar enfoques previos (definiciones o taxonomías) e intentar dar cuenta del rasgo o relación de ORGANIZACIÓN sin tomarlo como dado. Se trata, pues, de esclarecer la noción de nivel de organización por medio de una reconstrucción epistemológica del concepto de nivel.

El análisis se ordena en torno a 5 rasgos fundamentales que defiendo que pueden abarcar cualquier sentido posible del término y se ordenan de acuerdo a su grado de generalidad. Estas características son *composición*, *integración*, *emergencia*, *control* y *organización* y representan, precisamente, diferentes tipos de relación entre grupos de elementos, que pretendo articular en una perspectiva global. La reconstrucción procede de lo más general a lo más específico, empezando por un potencial mínimo común de todo concepto de nivel (composición) y avanzando gradualmente en la imposición de ulteriores condiciones restrictivas de esa generalidad hacia conceptos más específicos y operativos.

Para entender mejor la organización (biológica) podemos intentar considerar organización como un conjunto (*cluster*) de tipos de niveles diferentes, pero relacionados, cada uno con diferentes tipos de rasgos y características comunes a otros sistemas (o modos de ordenar cosas o fenómenos), pero que en conjunto despliegan los rasgos de la organización (biológica) plena, como en "niveles de organización ". En términos de la **jerarquía de especificación** de Salthe (1993) podemos presentar esos tipos diferentes de niveles de acuerdo a la especificación progresiva de su relación:

{composición {integración {emergencia {regulación {organización}}}}}

3.1. Composición

El primer paso requiere un cuidadoso análisis de lo que implica o no ser niveles de composición y así hay que considerar, al menos, 5 aspectos: *anidamiento,* relación de *orden parcial* como condición formal para cualquier concepto de nivel, cuál si algún sentido de *precedencia* es necesario, naturaleza *homogénea* frente a *heterogénea* de las partes componentes y el enigma *discreto/continuo*.

La condición básica de la relación de composición entre niveles es la más general, excluyendo únicamente la mera agregación en algún continente. En su generalidad, nivel de composición puede ser conceptualmente próxima a la noción de sub-sistema, pero se necesitan ulteriores especificaciones para clarificar su sentido.

a- *Anidamiento* (*nestedness*).

¿Debemos considerar sólo el caso *anidado* estándar? ¿Relación parte-todo como "hecho de tal materia"? Es decir ¿hemos de entender composición sólo como inclusión material? ¿O son también posibles otros conceptos composicionales más amplios, funcionales, dinámicos? Cómo encajar la composición en casos distintos como los de organismos - poblaciones, poblaciones - ecosistemas, o membranas - células, entre otros muchos. O cuando se trata de niveles de diferente composición superpuestos, pero en virtud de relación especifica que permita el uso de nivel con propiedad.

Si el concepto de nivel ha de ser aplicable también en estos casos problemáticos, la caracterización mínima de la relación de composición ha de ser lo bastante amplia para acomodarlos, sin hacer del anidamiento una condición obligatoria.

b. Relación de orden parcial.

En general, los niveles deben mantener una relación orden parcial entre ellos. Aparentemente esta condición puede resultar muy exigente como condición mínima (pues impone una modalidad de relación determinada) pero necesaria si se analiza cuidadosamente. Sin una relación de orden parcial la propia noción de nivel puede diluirse: niveles reversibles con respecto a la misma relación o varios niveles a un mismo nivel serían posibles.

c. Precedencia.

Algunos autores han propuesto la idea de precedencia como condición pero sólo sería necesario añadirla si se tomaran en consideración posiciones fuertemente holistas (aquéllas que atribuirían prioridad al todo). Es una condición que sólo se sostiene, si lo hace, en casos anidados y, en consecuencia, introduce una restricción innecesaria, dado que en casos no-anidados la precedencia como tal puede estar ausente. Por lo tanto, en general, añadir la relación de precedencia o no es válida como rasgo genérico o no añade nada a la relación de orden parcial.

d. Homogéneo vs. heterogéneo.

Es preciso abordar también la cuestión de si consideramos sólo componentes homogéneos (indiferenciados) o hemos de incluir la posibilidad de componentes heterogéneos (diferenciados). A este respecto, para los enfoques organizacionales (así como para los mecanicistas), es importante señalar que tratamos con partes diferenciadas, no simplemente con componentes equivalentes. Así, en general, la relación de composición debería ser capaz de acomodar ambos casos, que serían distinguidos subsecuentemente de acuerdo a ulteriores especificaciones de las relaciones.

e. Continuo/discreto

Por último, aunque sólo heurísticamente y se admita que surgen en un trasfondo de continuidad y grado, se requiere que los niveles sean en algún sentido discretos

3.2. Integración.

Esta relación responde a la cuestión de cómo están las partes constituidas en todos. Es una relación que especifica un tipo particular de relación de composición: aquella que produce un nivel superior coherente (en términos de Collier: *cohesivo*). Esta relación da lugar a un tipo de unidad autocontenida caracterizada por sus propias propiedades y por una extrema integración de sus partes (aquí se incluyen desde sistemas auto-organizados básicos o sistemas disipativos alejados del equilibrio, hasta organismos).

3.3. Emergencia.

Si la integración especifica la relación de los todos a sus partes, la relación de emergencia especifica más allá el proceso particular de formación de (al menos) algunos de esos todos. Clásicamente, la emergencia implica las dos características de **impredictibilidad** y **novedad cualitativa**. Aunque se suele recordar menos, también implica clásicamente el enraizamiento del nivel superior en el nivel inferior, está "*grounded*" en el nivel inferior.

3.4. Regulación (control).

Se refiere a la relación por la cual los niveles superiores, derivados en algún modo (por medio de la integración y la emergencia) de los niveles inferiores, ejercen a su vez alguna influencia sobre esos mismos niveles inferiores. Esta relación entre niveles debería dar cuenta tanto de las formas más discutidas de *causación descendente* como de otros casos aparentemente más simples como, por ejemplo, los de regulación cibernética (o control).

3.5. Organización.

La relación de organización entre niveles implica la combinación de los cuatro rasgos precedentes, todos ellos presentes. Por lo tanto, al hablar de organización, nos estamos refiriendo a sistemas complejos que implican una forma de interrelación entre elementos que va más allá de la mera composición. Nos referimos a sistemas que presentan propiedades globales integradas, propiedades que, si bien se derivan de las de sus partes componentes, son emergentes y, además, sistemas que regulan de alguna manera el comportamiento (dinámica) de sus elementos constitutivos.

Tradicionalmente se ha mantenido en biología que la *organización* es el rasgo fundamental de los seres vivos. Sin embargo, el concepto no pertenece de forma generalizada al vocabulario teórico de la biología contemporánea dominante aunque eso está cambiando.

> The organization of an entiy refers to the arrangement of its componente parts and their operations (functions) and to how they result in the capacities of the whole or the phenomena in which it appears. Often, organized entities are complex and hierarchical: their parts are themselves organized entities (Etxeberria & Umerez 2012).

En consecuencia, en este enfoque se plantea que la organización biológica exhibe precisamente esa relación internivélica más exigente y que, en cualquier caso, tanto los enfoques sistémicos como los mecanicistas en biología deberían ser capaces de dar cuenta de los niveles en términos de las relaciones internivélicas específicas que constituyen esa organización.

Para dotar de pleno sentido a esa relación, propongo acudir al concepto de **constricción** como materialización operacional clave de las relaciones de organización entre niveles:

> Constraint refers to a reduction on the degrees of freedom of the elements of a system exerted by some collection of elements, or to a limitation or bias on the variability or possibilities of change in the kind of such elements (Umerez & Mossio 2012).

Este uso del concepto se deriva de la propuesta inicial de Polanyi (1968) a partir de la idea de condición de contorno (boundary condition) y el control dual en la organización viviente, junto a su desarrollo posterior por Pattee (p.e.: 1968, 1972) quien introduce explícitamente en la Biología Teórica el concepto de constricción como herramienta explicativa derivada de la Mecánica, a partir de la distinción entre leyes y reglas y clasificación en holonómicas (estructurales) y no-holonómicas (funcionales).[2]

4. Algunas consecuencias del análisis conceptual

Con esas bases, planteo que la relación de composición o constitución no es suficiente para satisfacer las necesidades explicativas del enfoque mecanicista, que esa relación es también asimétrica al igual que la de causalidad y que, por lo tanto, no puede servir para descartar la causación descendente y, en definitiva, que el aspecto de organización no puede tomarse como dado sino que ha de justificarse explícitamente.

La mera composición no basta:

[2] Ver Umerez & Mossio 2012 para una primera aproximación a este tema que no se puede proseguir aquí.

A pesar de que en sus escritos algunos de sus defensores (como Bechtel o Craver) así lo hacen, podemos preguntarnos si los enfoques mecanicistas precisan restringirse a sí mismos (su concepto de *nivel de mecanismo*) a su sentido composicional y a continuación verse a obligados a cualificarlo con precisiones acerca de la *organización* (!) de los componentes.

Asimismo, resulta cuestionable su limitación a una concepción de la composición como simétrica para excluir la relación causal dado que de ser correcta la reconstrucción aquí presentada la composición es también a-simétrica (como la causación). La relación de composición entre niveles está parcialmente ordenada y por lo tanto es a-simétrica; el orden particular de cualquiera niveles puede ser revertido pero sólo de acuerdo con *otra* relación, no con la misma.

En este sentido, la posibilidad de causación descendente inter-nivel no puede ser excluida planteando que es una relación constitutiva. Po lo demás, sin querer obviar las dificultades de formular una concepción bien fundada de la causación descendente, ¿qué sentido tiene la causación *top-down* si no es internivelica / intrasistema?; ¿qué criterios (qué relaciones) justifican atribuciones *top-down* en otro caso?

Estos autores afirman repetidamente que, por ejemplo, "Mechanisms are organized collections of components and their activities" (Craver & Bechtel 2007) o que "… levels of mechanisms are componential levels (…) such haphazard parts are not parts in the same sense that …[mechanisms]" (Craver 2009). Y esta es precisamente la cuestión, ¿por qué? Se necesita algo más que la composición. La organización exacta de las partes (a diferencia de la sola extensionalidad mereológica) reclama la integración más allá de la mera composición.

No obstante, su insistencia en que "Mechanistically mediated effects are hybrids of constitutive and causal relations, where the constitutive relations are interlevel, and the causal relations are exclusively intralevel" (Craver & Bechtel 2007) o que "… all of interlevel relations are symmetrical" and "… many causal intuitions conflict with interlevel causes" (Craver 2009) sólo se puede entender porque están considerando únicamente un tipo de relación inter-nivélica: la de la simple composición anidada. El objetivo de este trabajo ha sido precisamente mostrar que para dar cuenta de la organización biológica hemos de recurrir a conceptos más complejos de nivel y de relación internivélica

Agradecimientos

El autor agradece la financiación del Ministerio de Economía y Competitivad y de la UE, Proyecto FFI2011-25665, y del Gobierno Vasco, Subvención a Grupos IT 505-10.

REFERENCIAS

Bechtel, W. (2008) *Mental Mechanisms*. London: Routledge.

Craver (2007) *Explaining the Brain*. Oxford: Clarendon Press.

Craver (2009) Levels of mechanisms: a field guide to the hierarchical structure of the World. In J. Symons & P. Calvo (eds.) *The Routledge Companion to Philosophy of Psychology*, London: Routledge, pp.: 387-399.

Craver & Bechtel (2007) Top-down causation without top-down causes. *Biology and Philosophy* 22: 547- 563.

Etxeberria, A. & Umerez, J (in press) Organization. In Dubitzky, W., Wolkenhauer, O., Cho, K-H. & Yokota, H. (eds.) *Encyclopedia of Systems Biology*. New York: Springer.

Pattee (1968) The physical basis of coding and reliability in biological evolution. In Waddington, C.H. (Ed.), *Towards a Theoretical Biology 1, Prolegomena*, Edinburgh University Press, Edinburgh, pp. 67–93.

Pattee (1972) Laws and constraints, symbols and languages. In: Waddington, C.H. (Ed.), *Towards a Theoretical Biology 4, Essays*, Edinburgh University Press, Edinburgh, pp. 248–258.

Polanyi (1968) Life's Irreducible Structure. *Science* 160, pp. 1308-1312.

Salthe (1985)

Salthe, S. (1993) *Development and Evolution*. Cambridge, MA: The MIT Press.

Umerez, J & Mossio, M. (in press) Constraint. In Dubitzky, W., Wolkenhauer, O., Cho, K-H. & Yokota, H. (eds.) *Encyclopedia of Systems Biology*. New York: Springer.

Wimsatt, W. (1976) Reductionism, levels of organization, and the mind-body problem. In G.G. Globus, G. Maxwell & I. Savodnik (eds.) *Conciousness and the Brain*, New York: Plenum, pp.: 199-267.

Wimsatt, W. (1994) The ontology of complex systems: Levels, perspectives, and causal tickets. In M. Matthen & R. Ware (eds.) Biology and Society: Reflections on Methodology. *Canadian Journal of Philosophy*, supp. vol. 20: 207-274.

REALISMO Y ANTIRREALISMO CIENTÍFICOS: EL CASO DE LAS LEYES

Presentación

En las últimas décadas, el debate entre realistas y antirrealistas científicos ha dado lugar a la reevaluación de antiguos argumentos y al surgimiento de una serie de propuestas alternativas. Los modernos realistas científicos encarnan una significativa reacción frente a la proliferación de actitudes escépticas o relativistas que habían alcanzado su máxima expresión en el período que marcó el ocaso de la filosofía de la ciencia de la primera mitad del siglo XX. Dentro del panorama actual encontramos una pluralidad de posiciones que defienden diversos tipos de realismo (metafísico, semántico, epistémico) y expresan diferentes combinaciones de estas alternativas. Pero, por su parte, el antirrealismo no ha dejado de tener adeptos; van Fraassen, por ejemplo, ha sorteado los desafíos escépticos y radicalmente relativistas elaborando argumentos que proporcionan una renovada versión del empirismo y han convertido a su autor en uno de los filósofos más influyente de los últimos tiempos.

Estos debates dan lugar a una serie de cuestiones relevantes para la determinación de las características del conocimiento científico. Entre ellas reaparece el problema del *status* de las leyes. Mientras los realistas generalmente se han inclinado a dar por supuesta su existencia, los antirrealistas científicos parecerían verse dispensados de encarar el problema de las leyes. Sin embargo, estos últimos no han podido evitar el reconocimiento de que ciertas proposiciones ocupan un lugar especial en el conjunto de las hipótesis científicas que las diferencia, por caso, de las llamadas generalizaciones accidentales. Esta cuestión, que ya había sido considerada por los filósofos de la ciencia de otras épocas, vuelve a plantearse en nuestros días con una particularidad: algunos autores niegan la existencia de leyes naturales aun cuando adhieren a ciertos principios fundamentales del realismo científico. En el presente simposio nos ocupamos de las tesis propuestas por dos exponentes de esta última perspectiva: Nancy Cartwright y Stephen Mumford. Nuestro propósito es señalar las debilidades que afectan sus argumentos. Las objeciones que presentamos a las ideas de Cartwright y Mumford no implican, sin embargo, asumir la defensa del realismo

nomológico. Adoptamos, por lo contrario, un punto de vista que rescata la función epistémica y el valor pragmático de las leyes sin comprometerse con la existencia de propiedades modales, esencias, universales, o cualquier otra clase de entidades semejantes.

SCIENTIFIC REALISM AND NOMOLOGICAL ANTIREALISM IN NANCY CARTWRIGHT

SUSANA LUCERO
Universidad de Buenos Aires

ABSTRACT: The present communication performs a critical analysis of the main thesis that Nancy Cartwright supports related with the laws status, its explanatory function and the sources which they come from. It also explores thoroughly the relevance of the arguments supporting the thesis that Cartwright defends. Like other authors (George Molnar, Sphen Mumford) her proposal is included in the nomological antirealism, that is the conception which claims that laws lack an independent existence from our methodological procedures and cognitive processes that ultimately make the laws to emerge. It also is in favor of the existence of causal powers or capacities in nature, thereby assuming a realistic postulate. This paper attempts to show a conflict between the quasi-instrumental function assigned to the laws and their descriptive character of the properties and causal sequences. We believe that Cartwright's nomological antirrealism is not consistently inserted in her version of the scientific realism of entities.
KEY WORDS: laws – capacities – nomological machines – dappled world

Introducción

En *Cómo mienten las leyes de la Física* (1983) Nancy Carwright presenta una concepción filosófica en la cual confluyen el antirrealismo nomológico con la tesis de la existencia de entidades teóricas, es decir con el supuesto ontológico del realismo científico. El puente que establece la relación entre las dos doctrinas es una teoría acerca de la explicación que comprende una crítica al modelo de cobertura legal y la defensa de un tipo particular de explicación causal basada en la inferencia a la mejor explicación. El desarrollo de estas ideas deja planteados dos tópicos centrales que Cartwright retomará en obras posteriores: el problema de las leyes y la postulación de poderes causales o capacidades. Nuestro análisis de la propuesta de Cartwright se dirige a señalar un conflicto entre la función cuasi instrumental asignada a las leyes y, por otra parte, su carácter descriptivo de propiedades y secuencias causales. Su concepto de ley muestra ser, en muchas ocasiones, ambivalente, pues oscila entre la idea de producto artificialmente logrado gracias a la intervención de las máquinas nomológicas y la descripción de la *naturaleza* de las cosas. En otras palabras, creemos que su antirrealismo nomológico no se inserta coherentemente en su versión del realismo científico de entidades.

1. Leyes y explicaciones

El centro de las reflexiones de Cartwright son las teorías físicas, aunque sus conclusiones pueden extenderse a todos los campos de investigación, inclusive a la Economía. Su punto de partida es la afirmación de que "los filósofos generalmente creen en las leyes y niegan las causas", pero Cartwright piensa que las leyes científicas en realidad "mienten". Las razones alegadas se apoyan en la comprobación de que su cumplimiento es parcial, limitado y aproximado. Las leyes de la física son literalmente falsas porque sólo se cumplen cuando se ponen en conjunción con cláusulas *ceteris paribus* de las cuales no tenemos certeza. Otro indicio que revela su insuficiencia es la ausencia de leyes que se refieran a las intersecciones de los distintos campos de fenómenos. Además se ajustan de manera más o menos aproximada a situaciones concretas, pero nunca están completamente libres de excepciones. Lo que en realidad está cuestionando Cartwright es el carácter universal de las leyes y de allí deriva su falsedad, aunque tal derivación no está fundamentada. De todos modos sus consideraciones parecen conducir a la siguiente conclusión: puesto que las leyes de la física mienten entonces no hay razones para creer en la existencia objetiva de las leyes, a lo sumo les deberíamos reconocer un valor epistémico y una función pragmática. Pero esta conclusión sería apresurada a la luz de reflexiones posteriores de la autora.

El escepticismo nomológico manifestado no lleva consigo un escepticismo acerca de las explicaciones. Más bien reivindica el papel de las explicaciones en la ciencia porque cree que la tarea de explicar científicamente tiene sentido y está convalidada por la práctica. Seguramente Cartwright considera que las leyes pueden, de todos modos, conservar un rol explicativo en un sentido instrumental. Analiza entonces dos modelos de explicación que corresponden a dos modos diferentes de "explicar". El primer sentido está asociado a la explicación de las leyes de bajo nivel o regularidades empíricas, también llamadas "leyes experimentales" o "fenomenológicas". Estas se deducen de los principios teóricos fundamentales de una teoría que son expresados comúnmente en forma de ecuaciones. Los partidarios del realismo científico duro sostienen que los principios implican las leyes experimentales y en esta medida las explican, por cuanto se trata de principios verdaderos. Nancy Cartwright, en cambio, opta por una posición más débil: reconoce el papel sistematizador y clasificatorio de las leyes en el seno de las teorías en un sentido análogo al que les había otorgado Pierre Duhem, aunque con una importante diferencia: para Duhem los principios teóricos no son ni verdaderos ni falsos, carecen de valor veritativo; Cartwright nos dice, en cambio, que son falsos puesto que "mienten".

El segundo sentido de explicar está representado por el modelo nomológico-deductivo de explicación de hechos particulares (modelo hempeliano), el cual prescribe la presencia necesaria de leyes que se asumen como verdaderas o aproximadamente verdaderas. Cartwright está dispuesta a

aceptar ambos tipos de explicación siempre y cuando se deje de lado la relación de las leyes con la verdad, conclusión que se desprende de su ya repetida tesis acerca de la falsedad de todas las leyes científicas. Hecha esta salvedad, privilegia el primer sentido mencionado de "explicar" por cuanto el modelo hempeliano para sucesos particulares no puede evadir el fenómeno de la subdeterminación: para cualquier clase de fenómenos habrá siempre más de una explicación satisfactoria, y algunas son incompatibles entre sí. Se podría cuestionar cuál es la ventaja de continuar sosteniendo un tipo de explicación basado en leyes cuando se ha decretado que éstas son siempre falsas. Observemos que Cartwright, en un sentido estricto, no niega que las leyes posean valores de verdad, de otro modo, caería en una flagrante contradicción. Por el mismo motivo, tampoco puede decir que ignoramos el valor de verdad de esas leyes. Es decir, respecto de las leyes la autora no asume una posición que sea estrictamente instrumentalista ni tampoco propiamente realista, por eso la hemos denominado "cuasi-instrumentalismo".

Cabe preguntarse ahora si hay otro tipo de explicaciones que prescindan de las leyes y de su presunta verdad/falsedad. La autora responde afirmativamente: las auténticas explicaciones son las que permiten inferir la causa de un hecho a partir de los rasgos distintivos del efecto. Los ejemplos abundan en la vida cotidiana y en el campo científico y remiten tanto a entidades observables como inobservables. Por ejemplo, "las trazas observadas en la cámara de niebla son producidas por una partícula en movimiento (probablemente un electrón)". Si convalidamos esta clase de explicaciones entonces tenemos que admitir la realidad de las causas. Ahora bien, el razonamiento por el cual accedemos a las causas de un fenómeno es una inferencia a la mejor explicación que concluye en la causa más probable. De acuerdo con Cartwright, por medio de la experimentación tenemos evidencia de la realidad de una conexión causal, que nos revela los poderes causales inherentes a las propiedades y procesos de la naturaleza; con tal postulación la autora sienta un principio básico del realismo entitativo. Al mismo tiempo parece quedar justificada la aplicación de la inferencia a la mejor explicación como una forma legítima de razonamiento, pues una vez admitida la realidad de las causas es razonable pensar que la mejor explicación de un fenómeno es aquella que le atribuye la causa más probable. Sin embargo esta estrategia de fundamentación –típica de los argumentos realistas- es discutible en este caso porque suscita inmediatamente una pregunta: "por qué es correcto aplicar la inferencia a la mejor explicación en vistas a postular la causa más probable de un suceso mientras que es rechazada cuando se la usa para determinar la verdad o aproximación a la verdad de las leyes? Se advierte también cierta circularidad en su modo de argumentar. En efecto, la introducción de propiedades dotadas de poderes causales depende de que se haya asumido previamente que hay relaciones de causa-efecto objetivas en la naturaleza, lo cual es otro modo de decir que existen poderes causales.

Diez años más tarde, en el artículo "In Defense of 'This Wordly' Cusality: Comments on van Fraassen's Laws and Symmetry", Cartwright termina disolviendo el concepto de ley en favor del de causa. Las secuencias causales (*causings*) y las propiedades modales son prioritarias tanto ontológica como epistémicamente. Los poderes causales ocurren en la naturaleza por todas partes, entre procesos observables e inobservables, y deberían figurar en las estructuras no observacionales como en las subestructuras empíricas de los modelos teóricos de la ciencia. La relación de causalidad no debe verse como una especie de cemento no perceptible que une porciones de propiedades sensibles no modalizadas; más bien se halla implícita en las descripciones del mundo y es allí donde hay que buscarla. La relación causa-efecto es la que verdaderamente realiza el trabajo que tradicionalmente se había adjudicado a las leyes, no son las leyes las que "explican" y "necesitan" sino las causas:

> "[...] la causa necesita su efecto –lo hace ocurrir, lo produce, y la ocurrencia del efecto es explicada por la ocurrencia de la causa en la forma robusta y realística- no subjetiva ni dependiente del interés- que Hempel pedía"(Cartwright 1993, 428).

¿Cuál es entonces la función de las leyes? La autora responde "las leyes describen lo que las causas son capaces de hacer".

Comprobamos que hay una tensión entre las diferentes caracterizaciones de ley. Si las leyes son descriptivas de procesos causales en la naturaleza entonces son verdaderas o falsas según que la descripción sea correcta o no, aun cuando nosotros no lleguemos a saberlo. Pero esta posición está lejos del slogan de que las leyes de la física mienten, porque si mienten son decididamente falsas. Ya no sabemos si las leyes tienen una función clasificatoria cuasi instrumental, como hemos dicho más arriba, si tienen capacidad explicativa o si pueden ser eliminadas a favor de las relaciones causales, como sostienen otros filósofos (Mumford, 2004).

2. Máquinas nomológicas y capacidades

En *The Dappled World* (1999), Cartwright adopta una nueva y original perspectiva para abordar nuestro conocimiento de las leyes y del modo como funcionan en la ciencia experimental moderna. Las leyes son producidas en condiciones ideales de experimentación como resultado de la organización de un experimento de laboratorio, en el contexto del modelo de alguna teoría científica. En otras palabras, surgen de un montaje teórico- experimental denominado "máquina nomológica". Sólo en circunstancias tan particulares las leyes se cumplen sin excepciones y pueden reputarse como verdaderas. Pero tales condiciones no son fáciles de producir ni de reproducir, se necesita que el modelo teórico que actúa como soporte se adecue correctamente al sistema físico de referencia. Bajo esta nueva descripción, las leyes exhiben los más variados rasgos que hayamos

decidido atribuirles: pueden describir regularidades o asociaciones de propiedades ocurrentes, secuencias causales singulares, les podemos permitir que soporten condicionales contrafácticos e inclusive agregar el *provisso* de que las regularidades en cuestión deben suceder con *necesidad*. En resumen, Cartwright trae al ruedo conceptos formulados por casi todas las tradiciones filosóficas que rivalizan sobre el tema.

Es oportuno hacer notar que la introducción de las máquinas nomológicas se inscribe en el marco de lo que la filósofa asume como un objetivo de la ciencia. En la dualidad "representar e intervenir" de Hacking, Cartwright elige el segundo término: queremos conocer el mundo para modificarlo. Las máquinas nomológicas son herramientas para la predicción, la manipulación y el control de la realidad que nos rodea, lo que sólo es posible en virtud de las leyes que las máquinas mismas producen. En vistas de que surgen en condiciones artificiales excepcionales, las leyes son limitadas en su alcance, dispersas y sumamente variadas. La ciencia actual no nos suministra un sistema deductivamente cerrado de oraciones nomológicas (contra la concepción Mill, Ramsey, Lewis -MRL) sino una malla de tejido emparchado, una multiplicidad de retazos legaliformes que se validan en un mundo igualmente veteado (the *dappled* world): "Es por eso que la Física, a pesar de ser una poderosa herramienta para predecir y cambiar el mundo es una herramienta de limitada utilidad" (Cartwright 1999, 9). Estas expresiones remiten a un instrumentalismo filosófico de una clase muy particular pues involucran un armado casi ingenieril de la producción del conocimiento. Contra lo esperable, Cartwright declara sus simpatías con el realismo, pues sostiene que toma "seriamente la insistencia de los realistas en que cuando usamos nuestra ciencia para hacer predicciones precisas o para lograr resultados artificiales, debe haber "algo correcto" en las afirmaciones que hacemos y en las prácticas que empleamos" (Cartwright 1999, 9). La ambivalencia filosófica de Cartwright continúa al exponer otras ideas que la empujan hacia un realismo metafísico de corte aristotélico. En efecto, el conocimiento más básico acerca del mundo no lo proporcionan las leyes sino las capacidades o poderes ocultos que se despliegan cuando se pone en marcha un sistema experimental gracias a la operatividad de las máquinas nomológicas. Las capacidades que así se manifiestan deben ser vistas como *naturalezas* en un sentido aristotélico, aquello que hace que una propiedad se comporte de determinada manera. Aun cuando Cartwright deja en claro que su adscripción a las naturalezas no implica un compromiso con otros principios de la filosofía de Aristóteles como la sustancia, la individuación o las esencias, piensa que aquellas proporcionan un conocimiento más abarcativo y más útil que el conocimiento de las regularidades, aunque ambos son necesarios. Las capacidades son propiedades disposicionales que la ciencia actual busca identificar a través de sutiles y muy elaborados experimentos; por ejemplo un átomo en estado de excitación tiene la capacidad de emitir fotones:

> "No esperamos que las naturalezas que son fundamentales para la Física se presentarán directamente en el comportamiento regular o típico de los fenómenos observables. Se requiere un difícil y delicado contexto experimental para revelarlas" (Cartwright 1999, 81)

Y en el enfrentamiento entre el privilegio cognoscitivo que los empiristas otorgan a las propiedades observables y a las leyes como regularidades, por una parte, y las disposiciones que se hallan en la naturaleza de las cosas por otra, no duda en declarar su preferencia por estas últimas.

Las críticas que se han formulado a lo largo de este trabajo ponen de manifiesto las ambivalencias conceptuales en que incurren los argumentos de la autora y favorecen la conclusión de que su antirrealismo nomológico no se inserta coherentemente en la versión del realismo científico que propicia. Al parecer, Nancy Cartwright construye una posición ecléctica sobre estos importantes tópicos al intentar integrar –en mi opinión, con poco éxito— doctrinas fuertemente realistas con conceptos instrumentalistas que resultan finalmente irreconciliables. Tal vez su postura sea la mejor ejemplificación de ese mundo "veteado" que la imagen científica –de acuerdo con su propia metáfora— intenta representar.

REFERENCIAS

Bailer-Jones, D. 2001, "Review of The Dappled World", *Erkenntnis* 54,. 411–415.
Cartwright, Nancy. 1983, *How The Laws of Physics Lie*, New York, Oxford University Press.
Cartwright, N. 1993, "In Defense of this Worldly' Causality: Comments on van Fraassen's Laws and Symmetry" 1993, *Philosophy and Phenomenological Research*, vol. 53, N° 2, 423-429
Cartwright, N. 1999, *The Dappled World: A Study of the Boundaries of Science*, Cambridge, Cambridge University Press.
Cartwright, N. 2007, *Causal Powers: What are They? Why Do We Need Them? What Can Be Done With Them and What Cannot?*, Centre for Philosophy of Natural and Social Science (LSE).
Cartwright, N. 2007, "What Makes a Capacity a Disposition?" in *Causal Powers: What are They? Why Do We Need Them? What Can Be Done With Them and What Cannot?*, Centre for Philosophy of Natural and Social Science (LSE).
Mumford, S. 2004, *Laws in nature,* London & New York, Routledge.

STEPHEN MUMFORD: MODALIDAD Y LEGALIDAD

NÉLIDA GENTILE
Universidad de Buenos Aires

ABSTRACT: Stephen Mumford makes a reconstruction of the so-called "nomological argument" in favor of natural laws and then formulates a dilemma in order to refute that argument. The nomological argument takes the form of an inference to the best explanation which postulates the existence of necessary connections - which are identified with the laws of nature -as the foundation of the regularities or patterns in the world. Against the existence of laws in nature, Mumford defends a modal realism based on the dispositions or causal powers: there are necessary connections but there are no laws. This paper will show that the Mumford´s suggestion only amounts to rephrasing the debate: rather than refute the nomological argument he maintains its structure and simply modifies the conclusion. As a consequence, the darts that Mumford threw against nomological realism reach the target of his own modal realism.
KEY WORDS: laws- nomological realism- dispositions- causality

Introducción

En *Laws in Nature* (2004), Stephen Mumford defiende un realismo modal que evita todo compromiso metafísico con las leyes. A diferencia de lo que ocurre con los modelos necesitaristas en los que las nociones de conexión necesaria y ley están íntimamente ligadas, Mumford defiende una postura intermedia entre el *realismo nomológico* –al que identifica con posiciones como la teoría DTA (Armstrong, Tooley y Dretske) o el esencialismo de Brian Ellis- y la visión "deflacionaria" que abrazan los humeanos. Mumford defiende una particular concepción a la que denomina "*realismo sin leyes*" (*realism lawlessness*). En acuerdo con los realistas nomológicos y en contra de los humeanos sostiene que hay conexiones necesarias o hechos modales en el mundo; pero en oposición a los primeros y en coincidencia con los últimos refuta la tesis que afirma la existencia de leyes. Así, el realismo sin leyes que Mumford propone no puede considerarse una posición antirrealista como tampoco una posición enteramente contraria a la doctrina de Hume: son las conexiones necesarias y no las leyes lo que se requiere para explicar las regularidades (conjunciones constantes). Mumford propone una ontología en la que algunas de las principales cosas que componen el mundo, sus propiedades, ya poseen modalidad, tienen conexiones necesarias con otras propiedades, y cuando éstas se instancian en los objetos naturales del mundo hay necesidad natural.

A fin de fundamentar su propuesta, Mumford presenta la forma idealizada del argumento sobre el que, a su juicio, se apoyan las distintas versiones del realismo nomológico y contrapone un dilema para refutarlo. El argumento nomológico, como veremos más adelante, adopta la forma de

una inferencia a la mejor explicación que postula la existencia de conexiones necesarias –las que se identifican con las leyes de la naturaleza- como fundamento de las regularidades o patrones que hay en el mundo. En el presente trabajo se intenta mostrar que la propuesta de Mumford sólo intercambia los términos del debate: más que refutar el argumento nomológico mantiene su estructura y simplemente modifica la conclusión. Consideramos así que Mumford sólo ha elegido "la mejor" de las alternativas dentro de un "mal lote": ¿Por qué entonces no mantener una concepción regularista de las leyes en lugar de la necesidad?

1. Realismo nomológico

El realismo nomológico, como hemos adelantado, es una posición realista respecto de las leyes. Frente a la concepción humeana –una concepción deflacionaria de las leyes- que entiende las leyes como meras regularidades o patrones que pueden encontrarse en el mundo, el realismo nomológico las interpreta de manera más robusta, como una "*adición de ser"*.

En opinión de Mumford, las distintas versiones del realismo nomológico (las diferentes teorías acerca de las leyes que pueden agruparse bajo esta denominación) se apoyan, de un modo u otro, en el siguiente argumento que el autor denomina "argumento nomológico" (NA):

A. Hay un conjunto S de características/rasgos en el mundo.

B. Hay S porque hay leyes de naturaleza.

Los miembros de S son fenómenos del mundo tales como la regularidad, la necesidad, la contrafactualidad, la invariancia, la predictibilidad, y demás, que sólo son posibles porque hay leyes de naturaleza. Así, las leyes son tomadas como el fundamento ontológico de las regularidades que hay en el mundo. Nótese que el argumento adopta la forma de una inferencia a la mejor explicación que va desde la regularidad o el orden hacia la existencia de leyes de naturaleza reales como su explicación y fundamento.

A fin de ilustrar el realismo nomológico Mumford pasa a analizar la teoría de Armstrong[1]. En contra de la visión regularista, Armstrong considera que las regularidades no son necesariamente leyes e, inversamente, que las leyes no consisten en regularidades. La nota clave es que las leyes son concebidas como una relación entre universales: "Todos los F son G" es una ley en virtud de la relación de necesidad nómica N(F,G) entre las propiedades (universales) de ser F y de ser G. La relación entre F y G es una relación de *necesitación* natural, no una relación de necesidad lógica. Así, N(F,G) implica (x) ($Fx \supset Gx$) pero no a la inversa. Las leyes genuinas se manifiestan a través de regularidades pero no todas las regularidades expresan leyes: que cada uno de los F sea un G no

[1] La concepción necesitarista de las leyes es conocida como el modelo DTA (Armstrong-Tooley- Dretske).

implica que el universal F se encuentre siempre en una relación de necesitación con el universal G. Asimismo, en opinión de Armstrong, los universales no son trascendentes, no son *ante rem* sino *in re*, existen sólo en sus instanciaciones y, consecuentemente, no hay universales no instanciados. La relación de necesitación N(F,G) es en sí misma un universal de segundo orden instanciado en las leyes. En algunas ocasiones, la relación causal es observable, "percibimos causaciones", y ello es lo que nos permite establecer patrones de regularidad como "Todos los F son G". La única diferencia es que no se trata de una relación entre casos (*tokens*) sino de una relación entre tipos (*types*) postulada e introducida por su efectividad explicativa (Armstrong 1993).

De acuerdo con Armstrong, esta perspectiva tiene la ventaja, entre otras, de poder distinguir claramente entre genuinas leyes y meras uniformidades accidentales, cosa que no puede hacerse desde la visión regularista de las leyes. La diferencia entre una mera regularidad y una ley de la naturaleza reside en la relación entre los universales que los particulares instancian, esto es, en la relación de necesitación: N(F,G).

Mumford reconoce que la concepción de Armstrong tiene fundamentalmente dos virtudes: por un lado, elabora una crítica eficaz del enfoque humeano regularista y, por el otro, ofrece una elucidación unificada de las relaciones entre leyes, universales y causación. Sin embargo, considera que el argumento nomológico de Armstrong no resulta persuasivo y no brinda una justificación filosófica suficiente para un realismo acerca de las leyes. Menos aun si se tiene en cuenta que catorce años después de *What is a Law of Nature?* (1983), en *A World of Status of Affairs* (1993), el propio Armstrong parece haber invertido el argumento al afirmar: "La conexión causal entre tipos de estados de cosas sigue siendo un postulado, por supuesto. Pero quién ha pensado que las leyes no sean postulados, que no sean inferencias a partir de las regularidades en el mundo?" (Armstrong 1997, 229. Citado por Mumford). En ente último texto las leyes parecen ser meramente inferidas a partir de las regularidades, pero esa inferencia no es segura -afirma Mumford- ya que puede haber regularidades por alguna otra razón o, como piensan los humeanos, simplemente por ninguna.
Demostrar que hay leyes en la naturaleza, conforme al argumento nomológico, supone mostrar que hay algún rasgo de la realidad que no estaría allí si no hubiese leyes. Las leyes y solo las leyes deberían ser las únicas cosas que podrían fundar esos factores. Pero, de acuerdo con Mumford, la teoría DTA no consigue justificar adecuadamente B, esto es, que hay S porque hay leyes: otras cosas diferentes de las leyes podrían dar cuenta de S.

2. Dilema Central

Independientemente de estas limitaciones, Mumford enfrenta la teoría DTA –y podría aplicarse a cualquier otra variante del realismo nomológico- con el denominado "Dilema Central".

El realista nomológico sostiene que las leyes juegan un rol en tanto determinantes de la historia del mundo, una historia de eventos o distribución de propiedades. Pero si esto es así, el realista debe explicar cómo las leyes cumplen este rol. Se debe asumir, además, que las leyes son externas a aquello que determinan o gobiernan o bien que son internas a ello. He aquí entonces los dos cuernos del dilema (Mumford 2004, 158)

Primer cuerno:

Una teoría de las leyes externas debe ofrecer una explicación adecuada de cómo las leyes se relacionan con las cosas que gobiernan. La explicación más plausible, la DTA, aun tiene dificultades en explicar esta relación e implica una tesis increíble: esencialismo (*quidditism*).

Segundo cuerno:

Una teoría de las leyes internas debe ofrecer una explicación de cómo las leyes pueden estar adecuadamente relacionadas con las cosas que gobiernan. La candidata más plausible, una teoría reductiva de las leyes, resulta implausible para desempeñar su rol explicativo.

En otras palabras, podemos formular el dilema en los siguientes términos: o bien las leyes son externas a sus instancias o bien son internas. Si son externas a sus instancias, entonces no pueden gobernar o determinar sus instancias. Si son internas, no pueden gobernar o determinar sus instancias. Luego, en ningún caso las leyes pueden gobernar sus instancias.

En el caso de Armstrong, dado que la relación nómica es externa, entonces choca con el primer cuerno del dilema: si las leyes son externas a sus instancias, entonces no pueden gobernar o determinar sus instancias. Por otra parte, en opinión de Mumford, la teoría de las leyes de Armstrong implica *quidditismo* acerca de propiedades, pues una vez que se ha permitido que en otros mundos los universales puedan estar nómicamente relacionados con universales distintos de los de nuestro mundo, entonces para sostener que se trata de la misma propiedad F sólo queda postular una esencia o *quidditas.*

El segundo cuerno del dilema se aplica a las denominadas teorías reductivas acerca de las leyes, y Mumford analiza particularmente la concepción de Lowe que reduce las leyes a clases y sus atributos. Así, las leyes son construidas a partir de otros elementos y son entonces reducibles a ellos. La cuestión que el dilema genera es cómo pueden las leyes gobernar o jugar un rol determinante sobre aquellas cosas a las cuales se reducen. Este tipo de crítica ha sido formulada por el propio Armstrong en contra de la teoría regularista (bajo una interpretación no humeana): si las leyes se agotan en sus instancias, entonces no pueden cumplir el rol de explicarlas.

Dado que, según Mumford, las leyes no logran cumplir con el rol que el realista nomológico le asigna, entonces no hay razón para mantenerlas. Pero la metafísica sin leyes que el autor defiende no equivale a adoptar el regularismo humeano; por el contrario, Mumford cree que una concepción

alternativa podría explicar S del mismo modo –o aun mejor– de lo que lo hacen las leyes: hay otras cosas que pueden asumir ese papel.

3. Realismo modal

Conforme con la conclusión del dilema, en la medida en que las leyes no cumplen con su propósito, puede afirmarse entonces que no hay leyes en la naturaleza. Pero ello no significa, tal como sostienen los humeanos, que no haya necesidad en el mundo. Por el contrario, haciendo uso de la terminología introducida por Nancy Cartwright, Mumford afirma que la necesidad natural depende de la existencia de propiedades *modales*. En opinión de Mumford, la necesidad en la naturaleza es una *necesidad de re*[2]. Distingue así distintas fuentes de necesidad *de re* que se expresan en relaciones de inclusión, relaciones de exclusión y relaciones de producción.

Hay relaciones necesarias de inclusión que no dependen del significado o la forma lógica –no son analíticamente necesarias- sino de la naturaleza de las propiedades involucradas. Ser coloreado y tener forma, o tener forma y poseer tamaño son propiedades entre las cuales se da una conexión necesaria: toda superficie coloreada tiene una forma y no puede existir forma sin tamaño. Hay también relaciones de exclusión necesarias tales como ser totalmente rojo o ser totalmente azul. Y hay un tercer tipo de necesidad *de re* -la más interesante a juicio de Mumford-, presente en las disposiciones o poderes causales. Ser soluble y disolverse son ejemplos de este último tipo. A diferencia de las dos primeras clases de necesidad *de re*, las disposiciones o poderes causales tienen una naturaleza dinámica: son responsables de los cambios que se producen en los particulares. Pero las disposiciones o poderes causales no deben entenderse como propiedades de propiedades, esto es, propiedades de segundo orden que se agregan a las propiedades de primer orden. Mumford concibe las propiedades como conjuntos (*clusters*) de poderes cuyas identidades son por ello fijadas por extensión. Por esta razón, los roles causales de una propiedad le pertenecen necesariamente (Mumford 2004, 171).

De acuerdo con el realismo modal hay entonces conexiones necesarias en la naturaleza y esas conexiones son absolutamente internas a los poderes (disposiciones) involucrados; no son un elemento externo que se agregue a los poderes sino relaciones internas que existen sólo en virtud de la existencia de sus *relata* (Mumford 2004, 173). Así, un "*cluster* de poderes naturales" constituye un conjunto poderes internamente conectados e instanciados en nuestro mundo.

[2] Mumford también se refiere a la necesidad *de re* en términos de necesidad natural o necesidad metafísica.

4. Legalidad sin necesidad

Hemos visto cuál ha sido la estrategia de Mumford en contra del realismo nomológico: el Dilema Central se presenta, a su juicio, como el argumento más contundente en contra de la existencia de leyes en la naturaleza. Pero no es éste el punto más relevante sino otro, a saber, la intención del autor de mostrar que el espacio vacante que han dejado las leyes puede ser ocupado por otro tipo de entidades, las disposiciones o poderes causales conforme a su propuesta alternativa. Tal como sostiene Psillos, Mumford habla libremente del rol causal de las propiedades y reemplaza *role nómico* por *rol causal*[3]. Si esto es así -y parece serlo en virtud de la recurrente insistencia de Mumford de que la función de las leyes podría cumplirla alguna cosa diferente- pareciera que Mumford mantiene la estructura del argumento nomológico. Denominémoslo [NA_M].

A_M. Hay necesidad natural (un miembro del conjunto S_M).
B_M: Hay S_M porque hay poderes causales.

Ahora, siguiendo la idea van frasseeana en su crítica al argumento de la inferencia a la mejor explicación, podemos afirmar que Mumford eligió, a lo sumo, "la mejor de las alternativas" dentro de un mal lote. Y en virtud ello, cabría además preguntar si la concepción de los poderes causales no queda atrapada bajo el Dilema Central. Creemos decididamente que sí. Pero dado que el propio Mumford se ha adelantado a esta última objeción y ha ofrecido ya una respuesta, debemos entonces explicar por qué, a nuestro juicio, fracasa en su intento de refutar la objeción.

En virtud de que, según Mumford, los poderes causales no son un elemento externo sino relaciones internas que existen sólo en virtud de la existencia de sus *relata*, pareciera evitarse, en principio, el primer cuerno del dilema. No obstante, si ese fuera el caso, automáticamente su teoría se enfrenta al segundo cuerno que, en razón del reemplazo que Mumford establece entre el rol nómico y el rol causal, quedaría formulado como sigue: una teoría de las propiedades internas debe ofrecer una explicación de cómo las propiedades pueden estar adecuadamente relacionadas con las cosas que gobiernan. Y a fin de rechazar esta posibilidad, Mumford formuló la siguiente respuesta:

> Una relación interna es algo que existe cuando existen sus relata. La relación interna no es, por lo tanto, algún elemento extra que se agrega a los relata, tales como ocurre en el caso de la explicación de las leyes ofrecida por el realista ingenuo o en la versión más sofisticada de la teoría DTA. […] Las propiedades modales se autogobiernan (*self-governing*) (Mumford 2004, 197).

[3] Psillos, S. Stephen Mumford's *Laws in Nature*. Trabajo no publicado. Véase http://philpapers.org

Así, en la medida en que las propiedades se autogobiernan y no hay brecha alguna a través de la cual las propiedades tengan que ejercer su influencia, la dificultad del dilema, en opinión de Mumford, ya no se presenta.

Pero en este punto resulta apropiado traer a colación la observación de Psillos. Como hemos indicado, Mumford cree que las propiedades modales se autogobiernan, constituyen conexiones entre poderes que se agotan en esas propiedades. Sin embargo, Psillos subraya que la propiedad de causación no es una pura relación interna dado que implica por lo menos las relaciones espacio-temporales externas entre sus relata. En efecto, basta recordar el análisis clásico de la idea de causalidad para advertir que junto a la relación de conexión necesaria se requieren las relaciones de contigüidad espacial y continuidad temporal entre los relata. Así, conforme con el señalamiento de Psillos volvería entonces a surgir el primer cuerno del dilema y a pesar de que Mumford considere que son las conexiones necesarias y no las leyes lo que se requiere para explicar las regularidades (Mumford 2004, 30), su propuesta no logra desvanecer el dilema. De este modo, los propios dardos que utilizó en contra del realismo nomológico alcanzan el blanco de su realismo modal.

Tal como el propio Mumford reconoce, el Dilema Central ataca los mismos puntos débiles que van Fraassen encontró en la teoría realista de las leyes, el problema de la inferencia y el problema de la explicación (Mumford 2004, 197). En la medida en que, conforme argumentamos, la sofisticada metafísica de Mumford no resuelve las dificultades, por qué no seguir el reto van fraasseano en favor de una actitud ontológicamente más modesta: quizá sea mejor quedarse con la deflacionaria concepción humeana de las regularidades y abandonar los riesgosos senderos de la modalidad.

REFERENCIAS

Gentile, N. 2011, “Acerca de la solución de Armstrong al dilema de van Fraseen”, in *Epistemología e Historia de la Ciencia*. Córdoba. FFyH. UNC, vol. 17, Nº 17.
Mumford, S. 2004, *Laws in Nature*. London, Routledge.
Psillos, S. 2002, *Causation and Explanation*, Acumen & McGill-Queens, University Press.
Psillos, S. 1989, “Stephen Mumford’s *Laws in Nature*”. http://philpapers.org
Van Fraassen, B. 1989, *Laws and Symmetry*, Oxford, Oxford University Press.

LOS COMPROMISOS ONTOLÓGICOS DEL REALISMO SIN LEYES

RODOLFO GAETA
Universidad de Buenos Aires

ABSTRACT: In order to assess the cogency of the claims of nomological realists, Stephen Mumford makes use of the criterion of ontological commitment put forward by Quine. Mumford contends that scientists do quantify over scientific laws, when they talk about the three laws of Newton, by instance. So, natural laws are conceived as substances, according to Quine´s criterion. Mumford says that it is a mistake, in view that a substance exists by itself and could exist even if there were no other being in the world, whereas a law cannot exist in a vacuum. We believe that Munford,s arguments are wrong. For one thing, the fact that a number of people make use of existential quantifiers is not a sufficient condition for thinking that they are making a hypostasis. Using quantifiers, unless they were irreducible, could be just a way of talking, as Quine himself admits. Furthermore, in Quine´s doctrine, the criterion of ontological commitment and the ontological relativity go hand in hand. Therefore, the criterion is not intended to give a definitive answer to the question about what there is. On the contrary, Mumford contends that metaphysics, and only metaphysics, is able to do it. We conclude, then, that Mumford´s arguments fail to show that laws do not exist.
KEY WORDS: laws - existence - ontological commitment- nomological argument

En *Laws in nature* (2004), Stephen Mumford afirma la existencia de conexiones necesarias en la naturaleza pero niega que las leyes naturales existan. En el presente trabajo se examinan estas ideas y se argumenta que la posición de Mumford se apoya en un concepto inapropiado de existencia.

La discusión en torno a las relaciones que guardan los conceptos de conjunción constante, inducción y conexión necesaria, como así también el examen de las nociones de causalidad, de ley natural y de ley científica están estrechamente asociados a los análisis llevados a cabo por Hume, a punto tal que es corriente continuar clasificando las distintas posiciones sobre estos temas en términos de concepciones humeanas o no humeanas. Conforme a esta distinción, las concepciones humeanas comparten una actitud empirista, rechazan la postulación de realidades metafísicas y, en consecuencia, se oponen al reconocimiento de conexiones necesarias o cualquier otro tipo de relaciones modales entre los fenómenos naturales. A su vez, esto implica que impugnan la afirmación de la existencia de relaciones causales, por cuanto la noción de causalidad suponeuna conexión necesaria entre la causa y el efecto.

Antes de adentrarnos en el objetivo específico de este trabajo quiero llamar la atención sobre el alcance del rechazo a las postulaciones metafísicas por parte de los “humeanos”, porque este

aspecto puede tener alguna incidencia en las consideraciones que haré respecto de la posición de Mumford. La palabra "rechazo" conlleva una ambigüedad. En un sentido fuerte, puede decirse que rechazar una afirmación equivale a afirmar su contradictoria; mientras que en un sentido débil, representa la actitud de no suscribirla ni negarla, así abstenerse puede considerarse la forma más débil del rechazo. La cuestión es relevante porque no hay unanimidad entre los estudiosos de la obra de Hume acerca de su actitud con respecto a las conexiones necesarias y, a la postre, sobre la causalidad.

La interpretación clásica de la doctrina humeana parece atribuirle un rechazo fuerte: "Hume was taken as the great denier of such necessary connection and the one who conclusively showed that there were no such thing to be found in nature" (Psillos 2002, 4). Sin embargo, aunque esta fuera una correcta lectura de la obra de Hume, no se terminan de despejar todas las dudas. En efecto, negar explícitamente que exista algo *en la naturaleza* no significa negarlo *tout court.* Un empirista consecuente, en la medida en que el empirismo incluya evitar pronunciarse sobre cuestiones metafísicas, subrayaría esa diferencia. Pero hay otras interpretaciones de las tesis de Hume; Galen Strawson, por ejemplo, afirma que Hume no negaba la existencia de la causación aunque argumentaba que no podemos saber nada acerca de su naturaleza. En resumen, en virtud de que la conexión necesaria es para Hume un componente esencial de la idea de causalidad, su posición acerca de las conexiones necesarias podría identificarse con alguna de las siguientes alternativas: *i)* No existen conexiones necesarias en la naturaleza (donde "naturaleza" se refiere al mundo de la experiencia) pero no se afirma ni se niega la existencia de tales conexiones más allá del mundo perceptible. *ii)* No existen conexiones necesarias ni en la naturaleza ni más allá de ella (porque sólo existe el mundo de la experiencia). *iii)* No existen las conexiones necesarias en la naturaleza, tal cómo ésta se presenta en la experiencia pero existen conexiones necesarias -sobre cuyas características no podemos saber nada- subyacentes a las conjunciones constantes que muestra la experiencia. Los "humeanos", entonces, serían quienes adhieren a las posiciones *i)* o *ii)* mientras que, si los partidarios del "nuevo Hume" tuvieran razón, el propio David Hume no sería "humeano" (Psillos 2002,53)

Mumford, atribuye a Hume la tesis *ii),* pues no vacila en adjudicarle una concepción metafísica derivada de sus convicciones epistemológicas: el único mundo que existe es el de la experiencia y, en consecuencia, no hay lugar para las conexiones necesarias. Al asignarle al empirismo una dimensión metafísica, Mumford favorece de antemano la defensa de su propia argumentación, pues se propone presentar una tesis estrictamente metafísica acerca de la inexistencia de leyes naturales. Si un empirista paradigmático no evitó asumir compromisos metafísicos, no se podría objetar que Mumford hiciera lo mismo. Pero estas consideraciones acerca

de la doctrina de Hume le reportan un beneficio adicional, porque Mumford puede marcar un punto de coincidencia aun más importante, le permite decir que tanto él como Hume niegan explícitamente la existencia de leyes en la naturaleza aun cuando la metafísica defendida por Mumford sea totalmente opuesta a la que le atribuye a Hume.

Sin embargo, no debemos apresurarnos a aceptar que el hecho de que Hume negara la existencia de las leyes en la naturaleza refuerza la posición de Mumford. En primer lugar, parece bastante evidente que el concepto humeano de naturaleza difiere crucialmente del que tiene en mente Mumford. Cuando Hume intenta determinar si algo existe en la naturaleza, se refiere, como ya se ha indicado, al mundo de la experiencia. Así, las bolas de billar y sus movimientos forman parte del mundo de la experiencia. Pero con la salvedad de que se trata de una forma imperfecta de expresarlo, porque podría considerarse que la afirmación de la existencia de los objetos físicos va más allá de las impresiones o datos sensibles. Mumford, en cambio, entiende por naturaleza algo mucho más amplio que los objetos de la experiencia, o quizá algo que ni siquiera incluye el mundo fenoménico, por momentos parece que identifica la naturaleza con una realidad totalmente trascendente.

Así, los límites y las relaciones entre la metafísica y la ciencia se tornan sumamente problemáticos, pese a que Mumford pretende mantener la distancia entre ambas. Por una parte, establece una distinción entre la ciencia y la metafísica paralela a la que diferencia sus respectivos objetos de estudio:

> "Science deals only with the phenomena in the sense that its subject matter is constituted by the things that are either actually or in-principle observable. By contrast, it is metaphysics, and only metaphysics, that tells us what ultimately exists" (Mumford 2004, 6).

Y agrega "Science declines to consider metaphysical entities, so it is also, in this sense, an incomplete account of the world (Mumford 2004, 11). Aunque parece admitir que una de las funciones de las ciencias naturales es establecer qué tipos de cosas existen -por ejemplo, cisnes negros-, la cuestión de la existencia es en última instancia metafísica:

> "Metaphysics asks what really exists. It may well conclude that the phenomena are real, or a reliable and truthful representation of what there is, but to maintain so would remain a distinctly philosophical claim that is outside the remit of science" (Mumford 2004, 6).

No hay ninguna garantía, pues, de que las cosas que componen el mundo fenoménico, el que la ciencia estudia, existan efectivamente. Pero si la metafísica es la única que puede resolver qué existe y qué no existe, cabe preguntarse cuál es el papel que Mumford reserva a la ciencia. Por cierto, aun cuando sostiene que la ciencia y la metafísica pueden enriquecerse mutuamente, no

queda claro que ello sea posible en un sentido ontológicamente relevante porque, como ya hemos señalado, cree que únicamente la metafísica está en condiciones de informarnos cómo está compuesta efectivamente la realidad.

Llegamos así a un punto crítico de la discusión: el alcance del concepto de existencia. Aunque no de una manera categórica, Mumford adopta el criterio de compromiso ontológico formulado por Quine: quien acepta cuantificar los miembros de un dominio, está postulando la existencia de esas entidades. Y como es posible cuantificar las leyes naturales, se estaría dando por cierto que existen. Conforme a la línea de interpretación seguida por Mumford, hablar de las tres leyes del movimiento como lo hacen los físicos equivale a concebir las leyes como substancias. Pero esta conclusión no le parece apropiada, porque una substancia es algo que podría existir aun cuando no hubiese nada más que ella en el mundo y se supone que esta posibilidad carece de sentido en el caso de una ley. Asimismo, Mumford rechaza inmediatamente la alternativa de que una ley pueda ser algún tipo de entidad colectiva -una colección de eventos, por ejemplo- sobre la base de una premisa que utiliza varias veces a lo largo de su libro: si una ley fuera una colección, no podría "gobernar" los miembros de ese conjunto.

Pero antes de analizar en qué sentido puede decirse que las leyes naturales gobiernan es conveniente detenernos en el problema del criterio del compromiso ontológico formulado por Quine.

> "Si atendemos a las variables ligadas en conexión con la ontología no es para saber lo que hay, sino para saber lo que una determinada observación o teoría *dice* que hay y éste es muy precisamente un problema de lenguaje, mientras la cuestión ¿qué hay? es de muy otro linaje" (Quine 1962, 43).

La propuesta de Quine parece óptima para revelar qué tipos de cosas se postulan, pues nada puede expresar más directamente la creencia en una entidad X que decir que existe X o hay X. Sin embargo, deben tomarse en cuenta otras consideraciones. Una es la inevitable relatividad de las afirmaciones de existencia con respecto a un marco conceptual (Quine 1962, 36). Esta condición puede tener consecuencias importantes para evaluar una postura eminentemente metafísica como la de Mumford, porque las decisiones ontológicas pasarían a depender de la previa elección de un marco conceptual y a diferencia de la situación de los empiristas, cuyas motivaciones epistemológicas privilegian determinadas perspectivas ontológicas -como el fenomenalismo o el fisicalismo- no queda muy claro cómo se ha producido la elección cuando se parte directamente de una metafísica ni cuál sería el alcance de sus afirmaciones ontológicas. De todos modos, en vista de que, en opinión de Mumford, la metafísica es la única que puede decir qué existe *realmente*,

parecería que considera la cuestión en términos absolutos y no quedaría lugar para la relatividad ontológica.

La segunda consideración se refiere al hecho de que incluso el compromiso ontológico que resulta de la aplicación del criterio de Quine puede ser, en ciertos casos, sólo aparente. En efecto, Quine sostiene que cuando se dice, por ejemplo, que algunas especies biológicas son cruzables, se adopta una ontología de entidades abstractas –las especies— sólo si el lenguaje no cuenta con un modo de parafrasear la oración que evite la cuantificación comprometedora (Quine1962, 40).

Ambas cuestiones, la relativización de la ontología con respecto al marco conceptual seleccionado y la distinción entre las afirmaciones que asumen efectivamente una ontología y las que sólo lo hacen de manera aparente conducen a una reflexión cuando se las toma conjuntamente. En los dos casos se trata fundamentalmente de escoger una manera de hablar, pero los alcances de una y otra elección parecen bastante diferentes. En el ejemplo de la reificación de las especies, se trata de una opción lingüística basada, en última instancia, en razones estilísticas, no de una auténtica hipóstasis. En cambio, en la medida en que el uso de los cuantificadores no se puede reducir a expresiones más primitivas, la adjudicación de un dominio de entidades como valores de las variables equivaldría a pronunciarse acerca de qué cosas pueblan el mundo. Mumford diría que ésa es, precisamente, la tarea indelegable de la metafísica. Pero mientras para él las verdades de la metafísica tienen un estatuto privilegiado, para Quine se trata de verdades triviales, que no necesitan ninguna justificación especial una vez que se ha adoptado un esquema conceptual.

> "Esto es, según creo, lo característico de la metafísica, o, por lo menos, de la parte de la metafísica llamada ontología. Quien considere verdadera una afirmación ontológica de esa rama tiene que considerarla trivialmente verdadera" (Quine1962, 36).

Por cierto, Quine no dice que las afirmaciones metafísicas sean triviales *per se,* sino que lo son a partir del esquema conceptual elegido. Asimismo, asimila la aceptación de una ontología a la de una teoría científica (Quine 1962, 44). Pero inmediatamente vincula la elección de una teoría, y por extensión, la preferencia de una ontología a la influencia de factores pragmáticos, como la simplicidad. Así, considera que tanto el esquema conceptual fenomenalista como el fisicalista, tienen sus propias ventajas y cualquiera de ellos puede ser reconocido como fundamental conforme al interés que en cada ocasión prevalezca. La matemática, la física y todas las ciencias deben comprometerse con ontologías y lo hacen a veces "hasta el cuello", dice Quine. Pero el compromiso no parece tan serio, después de todo, porque significa simplemente inclinarse por uno u otro mito, el que más convenga a los intereses y finalidades del momento.

Toda esta rememoración del célebre ensayo de Quine está motivada, obviamente, por el hecho de que el propio Mumford sugiere medir su propuesta ontológica, un mundo sin leyes, con la vara del criterio quineano. No me resulta claro con qué intención decidió Mumford esa estrategia, en vista de que no parece favorecerlo, a menos que procurara, de manera indirecta, cuestionar las ideas de Quine acerca de los problemas ontológicos. Y digo que la posición de Mumford no se ve beneficiada a la luz de las tesis de quineanas porque me parece que de acuerdo con ellas las cuestiones ontológicas, entendidas a la manera tradicional, como lo hace Mumford, pierden el privilegiado lugar del que gozaron durante tantos siglos. Se podría interpretar que el criterio de compromiso ontológico propuesto por Quine logra unificar el concepto de ser, en contraste con el célebre *dictum* aristotélico. Frente a la posibilidad de que el ser se predique de muchas maneras, Quine sostiene que ser es simplemente ser el valor de una variable, de manera que no existiría una forma de ser para los individuos y otra para los atributos de tales individuos, en la medida en que nuestro lenguaje permita usar cuantificadores tanto para los primeros como para los segundos. Pero decir que cierta entidad es el valor de una variable no nos informa nada interesante acerca del concepto de existencia, así como nada nos indica acerca de cómo está compuesta efectivamente la naturaleza o la realidad. En verdad, el recurso al que apela Quine constituye una manera de eludir esas preguntas. Ya hemos visto que las afirmaciones puramente ontológicas - "Hay tal o cual tipo de cosas"- resultan triviales una vez que, conciente o inconcientemente, se ha adoptado un marco conceptual. La trivialidad de la pregunta ontológica y su irrefutable respuesta, está ya adelantada en las frases que abren el ensayo de Quine.

Podría contraargumentarse, sin embargo, que con ese modo de encarar el tema, el problema no se ha disuelto sino que se ha enmascarado o desplazado a otro terreno, el de la elección de un marco conceptual. Esta idea sugiere que ciertas consideraciones, intuiciones previas o algo funcionalmente equivalente a estos factores –la percepción y el trato con el mundo, por ejemplo- conducen a los sujetos a inclinarse por una ontología y esta inclinación determina la construcción de un marco conceptual. Además, el marco conceptual podría modificarse, alterarse o cambiar sensiblemente en función de la necesidad de reconocer otros tipos de entidades. Si así fueran las cosas, la elección del marco conceptual no produciría los compromisos ontológicos sino a la inversa, las elecciones ontológicas condicionarían las preferencias en cuanto a marcos conceptuales. La cuestión ontológica sería irreductible, entonces, precisamente en este sentido, en su carácter fundamental.

Creo que un escenario semejante a éste conforma las presuposiciones de los argumentos de Mumford. Se esfuerza por demostrar que las leyes no existen en la naturaleza, esto es, que las leyes no existen. Apela, aunque no de manera totalmente explícita a una de las consideraciones hechas

más arriba acerca del criterio de compromiso ontológico de Quine, la posibilidad de establecer que cierto uso de un cuantificador no tenga un sentido literal y sea, por lo tanto, eliminable. Así descarta que la expresión "ley natural" se utilice en algún sentido metafórico y concluye que las leyes naturales no existen de ningún modo. Adjudica a Hume una actitud metafísica y sostiene que la errónea metafísica humeana conduce a sus adversarios – paradójicamente, puesto que Hume negaba la existencia de leyes en la naturaleza- a la postulación literal de la existencia de leyes naturales. Pero para que el esfuerzo de Mumford tenga sentido, es necesario concebir que alguien interprete las reflexiones de Hume como la negación metafísica de la existencia de leyes y plantee la existencia real de las leyes. Ello lo habilita para desplegar una serie de argumentos con el objeto de mostrar que todo lo que se ha hecho para avalar la existencia de leyes naturales está mal encaminado, ya sea porque se trata de regularidades en la naturaleza pero no de leyes; ya sea porque si gobiernan ciertas regularidades, no pueden ser parte de lo que gobiernan y entonces no están en la naturaleza; ya sea porque no pueden identificarse con proposiciones que describen el comportamiento del mundo, dado que las proposiciones no están en la naturaleza. Sólo así, si alguien está dispuesto a afirmar la existencia de leyes en la naturaleza, si tiene significado afimarlo, *ipso facto* tiene significado negarlo, en ambos casos en un sentido absoluto.

La extensa argumentación de Mumford para rechazar la existencia de leyes naturales contrasta con la simplicidad de la respuesta que se podría dar sobre la base de las tesis de Quine acerca del problema ontológico, claro está que dentro de los límites de la relatividad de la ontología con respecto al marco conceptual. Las leyes naturales no existirían (o, si se prefiere, no existirían en la naturaleza) a menos que formaran parte del dominio de valores de nuestras variables ligadas. Pero esta solución no conformaría a Mumford, porque "no existirían" quiere decir aquí "dentro de nuestro marco conceptual", y eso de ningún modo permite suponer que haya existencias independientes de los marcos conceptuales. Si las leyes existieran, no sería en la naturaleza a secas, así como si se negara su existencia tampoco se referiría a la naturaleza aparte de todo marco conceptual. Pero esto último, la inexistencia absoluta de las leyes es lo que las convicciones metafísicas de Mumford lo llevan a creer.

REFERENCIAS

Mumford, S. 2004, *Laws in Nature,* London , Routledge
Quine, W.V.O.1962, *Desde un punto de vista lógico*, Barcelona , Ariel.
Psillos, S. 2002, *Causation & Explanation,* Montreal, McGill Press.

SCIENCE, TECHNOLOGY AND SOCIETY

CIENCIA, TECNOLOGÍA Y SOCIEDAD

CARTOGRAFÍAS DEL CUERPO: BIOPOLITICAS DE LA CIENCIA Y LA TECNOLOGÍA I

Coordinación:

EULALIA PÉREZ SEDEÑO y MARÍA JESÚS SANTESMASES
Dpto. Ciencia, Tecnología Sociedad
IFS-CSIC

Este simposio presenta los resultados de un proyecto de investigación del mismo título en el que, durante tres años, se han analizado, desde una perspectiva de Ciencia, Tecnología y Sociedad, las ciencias y tecnologías del cuerpo, el estatuto que los cuerpos ocupan y desempeñan en las prácticas científicas. A través del estudio de un conjunto de tecnologías (terapéutico-reproductivas, de asignación y reasignación de sexo y de mejora y reforzamiento de los cuerpos) hemos examinaddo las representaciones de los cuerpos sexuados, y en particular los de las mujeres, el papel que desempeñan en la tarea científica, y como receptores y proveedores privilegiados de biomateriales para esas tecnologías, al tiempo que son, resultado de los procesos y relaciones promovidos por éstas.

Hoy en día el cuerpo se ha convertido en el centro empírico y teórico que se aborda desde diversas disciplinas como la sociología de la salud y la enfermedad la sociología de las emociones o del deporte, la antropología cultural o los estudios de ciencia, tecnología y sociedad. Sin embargo eso no siempre ha sido así. Por lo general se suele explicar la ausencia del cuerpo en la literatura por el legado de René Descartes y su distinción radical entre mente y cuerpo, entre lo mental y lo material, el alma y la naturaleza y por el privilegio que concedió a la primera parte de la dicotomía sobre la segunda. Según Descartes los hechos corpóreos pueden tener impacto en los procesos de pensamiento, pero el producto final (el pensamiento 'en sí') tendría valor en la medida en que estuviera descorporizado.

El cuerpo como sujeto de conocimiento es diferente al cuerpo como objeto de conocimiento. El cuerpo es el entorno primario que habitamos, la frontera física que nos identifica en singular, y frente a los demás cuerpos. Los cuerpos han sido abiertos, troceados, modificados, descritos y gobernados, de ellos se han extraido muestras, se ha logrado alargar su esperanza de vida y mejorar

su calidad. Las expectativas del conocimiento formal de los cuerpos no pueden ser más halagüeñas en el mundo contemporáneo, pero representar los cuerpos e intervenir sobre ellos conlleva prácticas médicas y tecnológicas son prácticas sociales de co-producción que intervienen en la medicalización de las ideas estéticas de los cuerpos, de la alimentación como vehículo de cura y prevención de enfermedades y como promotora del bienestar; e incluye negociaciones y consensos que construyen y reconstruyen permanentemente saberes y prácticas de clasificación, normalización y transformación.

Los procesos médicos y tecnológicos son prácticas sociales de co-producción que intervienen en la medicalización de las ideas estéticas de los cuerpos, de la alimentación como vehículo de cura y prevención de enfermedades y como promotora del bienestar. Nada de todo lo cual sería inevitable sino que al co-evolucionar cuerpos y artefactos, la materia tecnológca interviene en la identidad de género y viceversa, de forma que la relacion entre cuerpos y tecnologías está en permanente proceso de ajuste. Nuestro trabajo ha examinado algunas "tecnologías bio-médicas", agentes, prácticas y discursos en un contexto de progresiva medicalización de los problemas sociales, y donde más que tecnologías médicas o de la salud, tendríamos que hablar de "tecnologías de vida". La biomedicalización instaura un nuevo orden "biopolítico" que transforma los cuerpos y crea nuevas identidades individuales y colectivas.

NUEVOS ESCENARIOS EN LA GOBERNANZA DE LA CIENCIA DE LA SALUD: EXPLORANDO ESPACIOS EMERGENTES DE COMUNICACIÓN ENTRE EXPERTOS Y PACIENTES

ANA TOLEDO CHÁVARRI
Universidad Camilo José Cela

ABSTRACT: Current Science, Technology and Society approaches understand scientific innovations as coproduced throught the interactions between social and techno-scientific orders. The agents in this co-production are multiple, from researchers to citizens, media and stakeholders groups. They all participate in the governance of science and establishing areas of debate regarding re-evaluate the forms of knowledge production.
A review of Health Communication most cited articles presents very different scenario. overall scenario that deepens the gap between experts and patients and promotes medicalization of life. Public participation, legitimation of non scientific knowledge, and debates for including social and political arguments in health gobernance are nor relevant in the current research in Health Communication.
KEY WORDS: Health communication; patients; experts; medicalitation: new communication and information technologies.

Este es un análisis exploratorio de las nuevas tendencias de comunicación en el ámbito de la salud que parte de la revisión de los 25 artículos con más impacto en cuanto a número de citas sobre 'Health Comunication' publicados en los últimos 10 años y buscados a través de la Web of Knowledge. Esta revisión explora escenarios emergentes de la comunicación en el ámbito de la salud y en particular analizar cómo estos conciben la relación experto-paciente.

Partía de la hipótesis de que las tendencias comunicativas en salud se desarrollarían principalmente a través de los nuevos medios tecnológicos que proporciona la Web 2.0 y que estas herramientas tecnológicas facilitarían la participación ciudadana, la legitimación de saberes no científicos y de argumentaciones sociales, políticas y cívicas en el ámbito de la salud. Si bien el escenario encontrado es bastante diferente, no sólo no se tiende a una relación más horizontal entre los ciudadanos y el sistema médico si no que se profundiza en la medicalización de la vida.

El concepto de medicalización surgido de Zola teoriza sobre la extensión de la jurisdicción, autoridad y práctica médica hacia áreas cada vez más amplias de las vidas cotidianas de la gente (citado en Clarke et alt. 2006). En un primer momento de este proceso comenzado a mitad del siglo XX, la medicalización incidía en problemas sociales relacionados con la moralidad y el cuerpo que pasaban de ser mirados desde la ley a serlo desde la medicina (drogas, alcoholismo, sexualidad…) y

más tarde ha ido extendiéndose a otros ámbitos de vida cada vez más amplios como puede ser la alimentación o las concepciones sobre los estilos de vida.

En la literatura especializada revisada existe consenso al señalar la importancia de Internet en la transformación de la comunicación para la salud de las últimas décadas (más de la mitad de los artículos se centran en la red como medio de comunicación). Se señalan como factores importantes de cambio la apertura y facilidad de acceso a la información y los recursos sanitarios (Nelson, Kreps et al. 2004, Eysenbach 2008), la posibilidad de llegar a más gente (Noar, Benac et al. 2007, Nelson, Kreps et al. 2004), de forma más específica (Noar, Benac et al. 2007, Lee, Levy et al. 2002, Kreuter y Wray 2003, Neuhauser and Kreps 2003), efectiva (Kreuter y Wray 2003, Snyder, Hamilton et al. 2004, Kreuter, Sugg-Skinner et al. 2005) y a menor coste (Strecher 2007). Así mismos se constata el alto uso de las consultas de salud que consisten el tercer motivo de uso de Internet, especialmente para conocer más sobre los tratamientos, menos para prevención (Bennett y Glasgow 2009).

Algunos artículos apuntan a las posibilidades que se abren a la participación, colaboración, empoderamiento de los ciudadanos (Eysenbach 2008, Pignone, Ammerman et al. 2003, Murray, Burns et al. 2005, Strecher 2007). Si bien en general no se profundiza en las vías, características o límites de esta participación sino que se asume como evidente o potencial. Así,

> There is however also a broader idea behind Medicine 2.0 or 'second generation medicine': the notion that healthcare systems need to move away from hospital-based medicine, focus on promoting health, provide healthcare in people's own homes, and empower consumers to take responsibility for their own health (Eysenbach 2008)

Sin embargo, mayoritariamente se entiende que la comunicación para la salud es información producida dentro del sistema médico para ser diseminada hacia una audiencia que actúa exclusivamente como receptora de mensajes y no como activa en la búsqueda o productora de su propio conocimiento. Se trata por tanto de aumentar la confianza (Nelson, Kreps et al. 2004, Strecher 2007), mejorar la credibilidad (Wathen y Burkell 2002) y persuadir (Kreuter y McClure 2004, Dillard y Shen 2005) a los pacientes para que sigan los mensajes dados.

Los objetivos a conseguir más nombrados son aquellos que se dirigen a mejorar la eficacia de los mensajes a través de la mejora de las estrategias de comunicación y lograr con ello mayores cambios de conducta en las poblaciones. La *modificación de conducta*s, objetivo en 13 de los 25 artículos, propone dejar de fumar o beber alcohol, comer fruta y verdura, aumentar la actividad física o las mamografías o practicar comportamientos sexuales de protección. En ningún momento a

lo largo de estos 25 artículos los mensajes son cuestionados, se asumen como cajas negras (Latour) ajenos a las posibles controversias y debates que existen en torno a ellas.

Para mejorar la eficacia de las comunicaciones se propone especialmente los *tailored messages*, mensajes personalizados o a medida (tratado en siete de los artículos) definidos como 'any combination of strategies and information intended to reach one specific person, based on characteristics that are unique to that person, related to the outcome of interest, and derived from an individual assessment' (Noar, Benac et al. 2007).

A la hora de hacer mensajes a medida para la modificación de conductas aparecen señalados colectivos concretos mujeres (Jacobs, Karavolos et al. 2005, Kreuter, Sugg-Skinner et al. 2005), grupos multiétnicos (Jacobs, Karavolos et al. 2005, en referencia a no blancos) y de bajos ingresos (Kreuter, Sugg-Skinner et al. 2005). Así por ejemplo, un artículo señala 'women will continue to need routine cervical cancer screening and should continue to practice protective sexual behaviors (e.g., abstinence, monogamy, limiting the number of sex partners, and using condoms)'. (Friedman y Shepeard 2007)

En este contexto, la concepción de la persona aparece caracterizada de forma genérica como público, pacientes, consumidores y apunta hacia una concepción de la salud donde el individuo asume cada vez mayor responsabilidad dentro del enfoque del riesgo:

> Tobacco, poor diet and physical inactivity, alcohol consumption, microbial agents (such as influenza and pneumonia), toxic agents (e.g., air pollutants such as asbestos), motor vehicle accidents, firearm injuries, unsafe sexual behavior, and illicit drug use. (...) As these preventable risk factors are themselves behaviors, individuals may be able to add years to their lives as well as reduce substantial suffering if they are willing and able to make the health behavior changes necessary to potentially avoid chronic disease and premature death (Noar, Benac et al. 2007)

De esta manera, las personas están obligadas a informarse a sí mismo sobre sus enfermedades pero también sobre sus predisposiciones; está obligado a tomar los pasos apropiados, es decir a ajustar la dieta y estilo de vida a ellos (Dumit 2006) En este régimen del yo la salud pasa a recaer por tanto en los pacientes se transforma así en una meta individual, transformándose en una responsabilidad moral. El manejo de la enfermedad pasa por una implicación personal en aumentar el conocimiento, la autovigilancia, la prevención, el tratamiento de los riesgos así como el consumo apropiado de bienes y servicios médicos y de otras formas de cuidado (Clarke et alt. 2006). Para Callon y Rabeharisoa, esos comportamientos responsables se normalizan, convirtiéndose en rutinarios y esperados, de modo que quienes no actúan de la misma manera forman un nuevo tipo de persona problemática (2003).

Lo social aparece como un conjunto de individuos responsables cuya unión para por compartir, legitimar la información de la red a través de diversos tipos de plataformas de comunicación online:

> expert system (e.g., identifying current ability and barriers to tailor food cooking and shopping skills messages), user navigation (e.g., an information library with accurate, timely information regarding obesity and diabetes), collaborative filtering (e.g., a personalized recipe book based on recommendations of similar users), and human-to-human (e.g., an online support group) interactions in the same program. (Strecher 2007)

Así vemos que los principales productores del conocimiento continúan siendo los expertos, que sin embargo no aparecen como sujetos situados, se obvia nombrarlos, se les asume. Son los expertos quienes deciden qué comportamientos se han de cambiar y quién debe cambiarlos,

> An expert system attempts to create an educational experience using methods similar to a real-world clinical encounter. (...) Selection of tailoring factors is based on two criteria: what is most important (predictive of behavior and behavior change) and what is most changeable (Strecher 2007).

En esta literatura se mantiene estrictamente la separación experto-no experto deslegitimando el conocimiento producido más allá de los principales actores biomédicos que se tienen en cuenta a la hora de diseminar y legitimar a los expertos.

La participación de los pacientes que se cita se ciñe al uso de las redes sociales en las que se produce interacción se disemina y legitima la opinión experta:

> enables users to see what their peers or others with a predefined relationship ('friends', 'colleagues', 'fellow patients' etc.) are doing; enables automated selection of 'relevant' information (based on what peers are doing and reading on the Web); enables reputation and trust management, accountability and quality control, and fosters viral dissemination of information and applications (it is this 'viral marketing' aspect that makes Web 2.0 applications so attractive to venture capitalists and public health practitioners alike). (Eysenbach 2008)

Como conclusión podemos decir que la comunicación en el ámbito de la salud redunda, por tanto en un aumento de la medicalización en el sentido de control de la producción del conocimiento y la información desde la biomedicina, así como a través del intento de ganar eficacia en que la vida cotidiana de las personas, y especialmente de las mujeres, esté guiada por concepciones biomédicas. Se produce una paulatina transformación de los individuos en pacientes perpetuos puesto que siempre habrá conductas más saludables, modos de vida más adecuados para el sistema sanitario que además pasan a ser vistos cada vez en mayor medida como consumidores intensificándose la producción de productos y servicios relacionados con ellos, desde los diagnósticos, a los productos

alimenticios, los sistemas de control de dolencias crónicas, o la prevención de éstas (Clarke et alt. 2006, Dumit 2006).

Desde este punto de vista los escenarios emergentes en comunicación para la salud no apuestan de forma general por una transformación hacia una gobernanza de la ciencia más horizontal, si no más bien reproducen y fortalecen las actuales jerarquías del sistema sanitario profundizando en la medicalización.

REFERENCIAS

Bennett, G. G. and R. E. Glasgow (2009). The Delivery of Public Health Interventions via the Internet: Actualizing Their Potential. *Annual Review of Public Health*. 30: 273-292.

Callon, M. And Rabeharisoa, V. (2003) 'Research 'in the wild' and the shaping of new social identities', *Technology in Society* (25), 193-204.

Clarke, Adele; Shim, Janet; Mamo, Laura, Fosket, Jennifer Ruth y Fisman, Jennifer, R. (2003) 'Biomedicalization: Tecnoscientific Transformations of Health Illness ans U.S. Biomedicine', *American Sociological Review*, vol. 68 161-194.

Dillard, J. P. and L. J. Shen (2005). 'On the nature of reactance and its role in persuasive health communication', *Communication Monographs* 72(2): 144-168.

Dumit, Joseph (2006) 'Illnesses you have to fight to get: Facts as forces in uncertain, emergent illnesses' *Social Science & Medicine* 62: 577–590.

Eysenbach, G. (2008). 'Medicine 2.0: Social Networking, Collaboration, Participation, Apomediation, and Openness', *Journal of Medical Internet Research* 10(3).

Friedman, A. L. and H. Shepeard (2007). 'Exploring the knowledge, attitudes, beliefs, and communication preferences of the general public regarding HPV: Findings from CDC focus group research and implications for practice', *Health Education & Behavior* 34(3): 471-485.

Jacobs, E. A., K. Karavolos, et al. (2005). 'Limited English proficiency and breast and cervical cancer screening in a multiethnic population', *American Journal of Public Health* 95(8): 1410-1416.

Kreuter, M. W. and S. M. McClure (2004). 'The role of culture in health communication', *Annual Review of Public Health* 25: 439-455.

Kreuter, M. W., C. Sugg-Skinner, et al. (2005). 'Cultural tailoring for mammography and fruit and vegetable intake among low-income African-American women in urban public health centers', *Preventive Medicine* 41(1): 53-62.

Kreuter, M. W. and R. J. Wray (2003). 'Tailored and targeted health communication: Strategies for enhancing information relevance', *American Journal of Health Behavior* 27: S227-S232.

Lee, S. H., D. A. Levy, et al. (2002). 'Surveillance for waterborne-disease outbreaks--United States, 1999-2000', *Morbidity and mortality weekly report. Surveillance summaries (Washington, D.C. : 2002)* 51(8): 1-47.

Murray, E., J. Burns, et al. (2005). 'Interactive Health Communication Applications for people with chronic disease', *Cochrane Database of Systematic Reviews*(4).

Nelson, D. E., G. L. Kreps, et al. (2004). 'The Health Information National Trends Survey (HINTS): Development, design, and dissemination', *Journal of Health Communication* 9(5): 443-460.

Neuhauser, L. and G. L. Kreps (2003). 'Rethinking communication in the E-health era', *Journal of Health Psychology* 8(1): 7-23.

Noar, S. M., C. N. Benac, et al. (2007). 'Does tailoring matter? Meta-analytic review of tailored print health Behavior change interventions', *Psychological Bulletin* 133(4): 673-693.

Pignone, M. P., A. Ammerman, et al. (2003). 'Counseling to promote a healthy diet in adults - A summary of the evidence for the US Preventive Services Task Force', *American Journal of Preventive Medicine* 24(1): 75-92.

Snyder, L. B., M. A. Hamilton, et al. (2004). 'A meta-analysis of the effect of mediated health communication campaigns on behavior change in the United States', *Journal of Health Communication* 9: 71-96.

Strecher, V. (2007). Internet methods for delivering behavioral and health-related interventions (eHealth). *Annual Review of Clinical Psychology,* 3: 53-76.

Wathen, C. N. and J. Burkell (2002). 'Believe it or not: Factors influencing credibility on the Web', *Journal of the American Society for Information Science and Technology* 53(2): 134-144.

A VUELTAS CON LOS LÍMITES DE LA VIDA HUMANA. DE AUSENCIAS Y PRESENCIAS EN LAS NARRATIVAS SOBRE EL EMBRIÓN HUMANO

LORENA RUIZ MARCOS
(UCM)
CARMEN ROMERO BACHILLER
(UCM)

ABSTRACT: Where to draw the limits of what a life is and, more precisely, a "human life"? This is a question that sparks many debates around the definition of the beginning and the end of human life. Building on this question, focusing on the beginnings of human life, this paper addresses a range of sociomaterial configurations that shape the limits of human life. Through the lens of posthumanism, we analyze the human embryo as a liminal space, one in which different controversies on the origin of human life are articulated, revolving around three dimensions (morphological, spatial and temporal ones). Our approach defies linear narratives about the human and points instead to its precariousness, instability and dependency, all of which are usually excluded from those linear narratives of success. This paper does also stress the absences and the assemblages that make the emergence of the embryo possible, as elements that allow us to think its relation to the human in the framework of alternative ontologies.
KEY WORDS: embryo; pothumanism; human life

1. Introducción

De forma cada vez más recurrente, en los últimos años en el Estado español todo un cúmulo de discursos, imágenes y prácticas se vienen desplegando sobre el embrión humano. En el marco de múltiples controversias, como las que giran en torno al aborto, la investigación con células madre o el desarrollo de bebés medicamento, el embrión es situado en el centro del debate a través de la acción de diversos actores que pugnan por estabilizar los contornos de la vida humana en términos temporales (*cuándo* se puede decir que el embrión es una vida), en términos espaciales (*dónde* adquiere el embrión su humanidad) y en términos morfológicos (*qué forma* ha de adoptar el embrión para pasar a habitar el terreno de lo humano).

Los límites de la vida humana se convierten así en objeto de disputa, en un terreno minado en el que la cualidad performativa del lenguaje se hace patente: llamar al embrión "no nacido", "vida en formación" o "conjunto de células" tiene consecuencias reales en el mundo, influye en las vidas de personas que de una manera u otra se ven afectadas por la regulación legal y la percepción social en torno a decisiones como el aborto o la posibilidad de evitar determinadas malformaciones genéticas antes del nacimiento de un hijo. Pero no sólo los discursos y las representaciones forman

parte del contexto en el que se dirime la vida humana. Las prácticas materiales (congelación de embriones, donación de óvulos, análisis genético de los embriones) también están delimitando las formas en las que el embrión se erige en referente de la vida humana en espacios como la clínica y el laboratorio.

De esta manera, a través del ensamblaje de múltiples elementos (discursos, imágenes, técnicas, tecnologías, leyes), se configuran versiones específicas del embrión (el embrión como material de investigación biomédica, el embrión como objeto sometido a una fuerte regulación legal, el embrión como argumento para decidir continuar o interrumpir un embarazo). La demarcación de los límites de la vida humana, especialmente en términos temporales (cuándo comienza una vida y, de forma más concreta, cuándo comienza una vida a ser *humana*), atraviesa de forma permanente los procesos por los cuales el embrión emerge como una entidad precaria, inestable, sometida a diversos ejercicios de portavocía y abocada a habitar un espacio difuso entre la definición con pretensiones de objetividad (por ejemplo, la que se puede encontrar en manuales de medicina) y las interpretaciones morales e ideológicas (por ejemplo, los discursos pronunciados desde esferas religiosas).

Este trabajo rastrea las formas en que diversas asociaciones de elementos heterogéneos actúan para delinear los orígenes de la vida humana. Atendiendo a la emergencia del embrión como resultado de múltiples configuraciones sociomateriales, entendemos que la precariedad y la dependencia son elementos constitutivos del embrión e indagamos en ciertas narrativas que por el contrario se basan en la linealidad, la certeza y la esencia para conceptualizar lo humano.

2. La vida humana desde un enfoque posthumanista

Las ciencias sociales han tratado históricamente con las preguntas acerca de la humanidad de lo humano, en la búsqueda de criterios que permitan diferenciar lo humano de lo no-humano y adscribir diferentes estatutos a cada categoría de seres vivientes. La humanidad de lo humano ha sido históricamente sustentada sobre el privilegio que resulta de la diferencia, primero, entre lo "vivo" y lo "inerte" y, después, entre la "vida humana" y las "vidas no humanas" (Butler 2004).

Pero trazar la línea divisoria entre la vida humana y las vidas no humanas no es tan fácil y evidente como a simple vista pudiera parecer. El posthumanismo nos ofrece herramientas para constatar que no es posible efectuar un corte "limpio" entre lo humano y lo no-humano y que la inestabilidad permea ambas categorías, pues lo que nos encontramos son posiciones que obtienen reconocimiento como humanas (o son designadas como no-humanas) mediante configuraciones sociomateriales que habilitan y actualizan esas posiciones.

A través de su ensamblaje con lo no-humano, la humanidad adquiere entidad y accede a la agencialidad que se le presupone. Sin embargo, la asunción de que agencialidad e independencia son inseparables provoca la desaparición de los elementos no-humanos de la narrativa, que quedan situados en los márgenes como meras herramientas a la merced de una humanidad en la que reside no sólo la agencia sino también la autonomía y la volición. En este trabajo trataremos de mostrar que, por el contrario, los elementos no-humanos son "ausencias" constitutivas de lo humano. Adoptamos así una mirada posthumanista, en la que lo humano no es origen ni esencia sino resultado y efecto, un punto de llegada contingente sometido a continuas revisiones e instituido sobre la invisibilización de su ensamblaje necesario con lo no-humano.

3. El embrión como objeto sociotécnico

Para abordar una de las formas concretas en que el embrión es actuado en el marco de disputas en torno a los límites y estatuto de la vida humana, acudimos a una serie de imágenes que forman parte de la campaña que la Conferencia Episcopal Española lanzó a los medios de comunicación para mostrar su rechazo a la investigación con células embrionarias, cuya fecha data del año 2005. La imagen de una familia compuesta por una pareja heterosexual y tres hijos (un bebé varón, un niño y una niña) aparece enmarcada por el eslogan "Todos fuimos embriones". Bajo esta instantánea aparece una tira horizontal de siete imágenes de tamaño mucho más pequeño, en las que se representan diferentes momentos del desarrollo embrionario y fetal, la imagen de un recién nacido durmiendo y, por último, la misma familia que acabamos de describir clausura la secuencia. Es interesante destacar aquí cómo la vida humana es inscrita de esta manera en una estructura lineal, en un ejercicio de continuidad en el que el paso de un conjunto de células a un feto, posteriormente a un bebé, y finalmente a una familia, se presenta como aproblemático y exitoso en todo caso. Están ausentes de la narrativa elementos tales como la incertidumbre, la elección personal o los requerimientos que ha de cumplir el embrión (realizar su división celular e implantarse en el útero, ninguno de los cuales son procesos inmediatos).

Las dos imágenes que abren la secuencia que estamos analizando corresponden al embrión en sus fases iniciales, cuando está formado por un conjunto de células. Estas primeras imágenes representan icónicamente el potencial para producir vida que contienen las células, aluden directamente a narrativas de desarrollo y evolución cuyo "producto final" sería un organismo humano vivo. La decisión de situar como primera imagen de la secuencia un embrión compuesto por dos células aparece como una reivindicación de la existencia de vida humana desde el momento mismo de la concepción (aunque en el proceso de fecundación es una única célula la que resulta de

la fusión de ovulo y espermatozoide). Este embrión estaría así simbolizando una especie de garantía de vida, una vida "de cuya continuación se hace responsable a un 'nosotros' global" (Duden 1991/1993, citado en Franklin, Lury y Stacey 2000, 36). Las imágenes del conjunto de células nos invitan, o más bien nos obligan, a confrontar la responsabilidad hacia la vida que o bien puede partir de esas células o bien puede verse interrumpida si la división celular no es exitosa. Pero, ¿quién es ese "nosotros" global al que aludíamos anteriormente? ¿A quién están interpelando estas células?

La tercera instantánea de la tira de imágenes empleada por la Conferencia Episcopal en su campaña es la de un embrión en sus estadios finales, probablemente alrededor la octava semana de embarazo, aunque también podría tratarse ya de un feto. En cualquiera de los dos casos, lo que vemos es una figura que ya reconocemos como humana, debido a la existencia de un rostro y órganos tales como un ojo, una mano o los pies. La morfología se torna aquí central en la construcción de los referentes de lo humano: la cara, un ojo, las manos o los pies, elementos que además deben aparecer diferenciados, en su singularidad, y situados en el contenedor (la piel) de un único individuo.

Las tecnologías de visualización que crean imágenes del feto en tres y cuatro dimensiones y permiten ver por tanto cómo se mueve o el gesto de su rostro, actúan como poderosas herramientas en la adscripción de una determinada "actitud" y agencia al feto, presentándole así como un ente autónomo respecto a la madre (a pesar, paradójicamente, de la dependencia que tiene de ella) y constituyendo una serie de indicadores de lo humano que irían incluso más allá de lo morfológico, como la capacidad de mostrar ciertas emociones, de sonreír, etc.

El concepto de cuerpo está operando bajo la lectura de estas imágenes de una forma central a la hora de realizar una interpretación de las mismas. Para otorgar al embrión un lugar entre lo humano, la posibilidad de reconocer un cuerpo se vuelve clave. Y el rostro es uno de los criterios principales que empleamos para discernir la existencia de un cuerpo y de algo "humano" (no sólo su morfología, sino incluso su mera existencia[1]). Judith Butler, siguiendo la argumentación de Emmanuel Levinas, habla del rostro como aquello que nos permite "explicar el modo como somos

[1] A este respecto el cirujano Pedro Cavadas, quien realizó en 2009 el primer trasplante parcial de cara en España, afirma que "no tener cara significa no estar" y que "con el trasplante se le pone [al paciente] *una cara humana*" (énfasis añadido). Ver Ferrado, Mónica L. "Entrevista con Pedro Cavadas: 'No tener cara significa no estar'". *El País*, 23 de noviembre de 2008, Extraído el 8 de mayo de 2010, de http://www.elpais.com/articulo/sociedad/tener/cara/significa/estar/elpepusoc/20081123elpepisoc_5/Tes.

También es significativo cómo la posibilidad de visualizar el rostro en las ecografías adquiere un valor emocional y actúa como garante de la existencia "real" de la vida humana en el interior del cuerpo materno. Así, en una página web que ofrece un servicio de obtención de ecografías en tres y cuatro dimensiones se subraya lo siguiente: "en ocasiones la escasez de líquido, las manos delante de la cara y/o el cordón umbilical no facilitan la visualización del rostro, sin embargo si se logra ver, lo cual ocurre en el 70/80% de los casos, son imágenes *inolvidables* por su calidad y *realidad*" (énfasis añadido). Ver Ecografías 4D, Extraído el 5 de mayo de 2010, de http://www.ecografias4d.com

interpelados moralmente, el modo como el otro nos demanda moralmente –una apelación moral que no pedimos ni somos libres de rechazar–" (Butler 2004/2006, 166). Y responder activamente a esta demanda "significa despertarse a lo que es precario de otra vida o, más bien, a la precariedad de la vida misma. [...] Debe ser una comprensión de la precariedad del Otro. Esto es lo que vuelve al rostro parte de la esfera de la ética" (Butler 2004/2006, 169). ¿Qué nos demanda entonces el "rostro" del embrión? ¿Que lo reconozcamos como una vida humana?

En el cuarto y quinto lugar de la secuencia nos encontramos con imágenes del periodo fetal, en las que el cuerpo de la madre está ausente por completo. De esta forma, la individualidad está operando sobre la percepción de lo que cuenta como una forma humana, y ello en dos aspectos. En primer lugar, las imágenes del feto separan a la mujer embarazada y al propio feto, que pasa a ser otro paciente más, un objeto clínico susceptible de ser observado y tratado (Price 1990, citado en Strathern 1992). La independencia del feto hace que se constituya en el referente de la vida, en el origen y potencial para generar vida, mientras que la madre aparece relegada a ser un mero entorno y recurso, "una tecnología 'habilitante' o […] un sistema de apoyo" (Strathern 1992, 49).

Un bebé recién nacido, durmiendo, del cual vemos sólo su rostro y una parte de un brazo, es la penúltima imagen de la secuencia. Con esta instantánea se consigue el efecto de linealidad perseguido: la secuencia "células-embrión-feto-bebé" alcanza así su objetivo final y muestra el éxito de la promesa que contenían las primeras células. Quizá sea este el motivo por el que las imágenes de las células y del embrión nunca se suelas presentar aisladas, sino acompañando a las del feto o del bebé. El cuerpo del bebé logra también elevar el tono emocional de la interpelación de esta secuencia de la vida humana, pues ahora nuestra mirada se encuentra frente a un rostro ante el que sentimos una mayor cercanía que la que experimentamos ante la imagen del conjunto de células. El bebé es alguien que apela a nuestros afectos y que podemos ya reconocer como humano, un rostro que, siguiendo a Butler y Levinas, muestra una vulnerabilidad extrema (es un bebé) y apela a nuestro reconocimiento solidario (en tanto que humano).

Pero todavía hay un último movimiento de clausura en la secuencia que estamos analizando: el bebé recién nacido aparece en la última imagen como parte de una familia (blanca, de clase media, heterosexual, acomodada, sana), la misma familia que ocupa el centro de los carteles de la campaña. Si todos fuimos embriones, como reza el eslogan, entonces todos somos nacidos, todos somos humanos y, en tanto que bebés, miembros de una (determinada) familia. La continuidad en términos genéticos, ideológicos o físicos entre padres e hijos se presenta como autoevidente en un marco narrativo lineal cuyos puntos de origen y llegada son ahora las células y la familia, respectivamente: "células-embrión-feto-bebé-familia".

Conclusiones

En este trabajo hemos entendido el embrión como un objeto sociotécnico, lejos de concepciones biologicistas o de meras construcciones médicas, es decir, un objeto que es actuado a través de un tupido entramado de actores, prácticas y significados que producen su materialidad y que hacen que ocupe un lugar central en las disputas por la definición y estatuto de lo humano.

Los desarrollos tecnológicos en el ámbito de la investigación y la intervención biomédica hacen que cada vez sea más complicado establecer la división entre lo humano y lo no-humano. En este escenario, la respuesta a la pregunta sobre la humanidad del embrión se torna cada vez más compleja y con mayores consecuencias sobre los horizontes de posibilidad que se abren tras una respuesta positiva (o negativa) sobre la humanidad del embrión. Si acordamos que el embrión no es una vida humana, ¿cuáles son las actuaciones legítimas sobre él? ¿Qué espacios puede pasar a habitar? Y si el embrión es visto por el contrario como una vida humana, ¿es necesario reconfigurar los límites entre ciencia y sociedad? ¿Qué formas de gobierno de la vida humana están siendo trazadas?

REFERENCIAS

Butler, J. 2004, *Undoing gender*, Londres y Nueva York, Routledge.

Butler, J. 2004/2006, *Vida precaria. El poder del duelo y la violencia*, Buenos Aires, Paidós.

Duden, B. 1991/1993, *Disembodying women. Perspectives on pregnancy and the unborn*, Cambridge, Harvard University Press.

Ferrado, M.L. 2008, 23 de noviembre, 'Entrevista con Pedro Cavadas: 'No tener cara significa no estar', *El País,* Extraído el 8 de mayo de 2010, de http://www.elpais.com/articulo/sociedad/tener/cara/significa/estar/elpepusoc/20081123elpepisoc_5/Tes

Franklin, S., Lury, C. y Stacey, J. 2000, *Global Nature, Global Culture*, Londres, SAGE Publications.

Price, F. 1990, 'The management of uncertainty in obstetric practice: ultrasonography, in vitro fertilisation and embryo transfer', en McNeil, M., Varcoe, I. y Yearley, S. (eds.), *The New Reproductive Technologies*, Londres, Macmillan, 123-153.

Strathern, M. 1992, *After nature: English kinship in the late twentieth century*, Cambridge, Cambridge University Press.

REPRESENTACIONES CROMOSÓMICAS: CUERPOS, CARAS Y MUESTRAS AMNIÓTICAS

MARÍA JESÚS SANTESMASES
Instituto de Filosofía, CCHS-CSIC

ABSTRACT: this presentation aims as drawing a trajectory of practices and ways of constructing a new epistemology of bodies, of their identities and representations by way of the description of chromosomes. From sex determination and Down syndrome diagnosis in the clinic, I show an origin of medical genetics in the practices of amniocentesis that contributes to accounting for a genealogy of prenatal diagnostics.
KEYWORDS: Chromosomes, bodies, the clinic and the laboratory, Down syndrome, sex determination, the gender and two sexes.

Los experimentos en las muestras extraídas de cuerpos humanos, de los vivos y de los cadáveres de las salas de anatomía, han generado identidad y representación. Entre la observación a simple vista y la mirada entrenada al microscopio se extienden todos los métodos de observación al cuerpo experimental, como lo ha llamado Ilana Löwy. La citogenética inventó en la década de 1950 los métodos con los que era posible observar y caracterizar los cromosomas -orgánulos del núcleo celular- que se dejaban teñir para ser observados al microscopio, a partir de muestras de médula ósea primero y a partir de 1960 con muestras de sangre periférica. Desde entonces el procedimiento de representación cromosómica de los cuerpos ha hablado de desórdenes –infertilidad y síndrome de Down entre ellos fueron los primeros entre 1959 y 1960.

Se presenta aquí un recorrido breve por la trayectoria de representación cromosómica del sexo y del síndrome de Down en el que los cromosomas hablarían de los cuerpos, de las facciones y las aptitudes mentales, como hablarían de fetos y embriones cuyas células empezaron a estudiarse en muestras de líquido amniótico procedente del abdomen.

La biologización cultural de los cuerpos les ha restado identidad. O quizá ha reconstruido la identidad en forma de representación *in vitro* de eso fenómenos que son el sexo y el retraso mental. Los cuerpos resultaron así cartografiados con criterios asignados a nuevos experimentos que acarrearon categorías epistémicas nuevas. La epistemología experimental se combina en este texto con el uso del género como categoría analítica para dar cuenta de los cruces e interferencias en la reconstrucción de las identidades de género y de aquellas asignadas a grado de normalidad y patología mentales.

En la mesa de laboratorio las células quedaron aisladas, sometidas a la tecnología de cultivo experimental para que su multiplicación diera cuenta de la salud o de la enfermedad. Esa genealogía tecnológica de las muestras, originada en los cuerpos, generó culturas subalternas de género y materia médica propia de las culturas androgénicas. Aunque transparente, los cuerpos de las

mujeres permanecían como objeto de estudio e inscripción ontológica en las nuevas tecnologías y en sus rituales analíticos y clasificatorios.

Propuesta para una cronología

En 1956 Tjio y Levan publicaron un famoso artículo en la revista sueca Hereditas en el que mostraban un nuevo número de cromosomas humanos: 46 pasó a reemplazar así el viejo número 48. Aunque cautos, habían podido contar 46 cromosomas en más de 265 experimentos con muestras de tejido de embriones procedentes de abortos legales (Martin, Kottler, Lindee).

Desde su laboratorio en las instalaciones atómicas británicas dedicados a estudiar los efectos biológicos de las radiaciones, Charles Ford y John Hamerton reprodujeron de inmediato los resultados de Tjio y Levan. Según ellos, fue clave no tanto el microscopio, con su capacidad amplificadora de aquellos orgánulos subcelulares diminutos, sino la técnica preparativa que hizo a los cromosomas visibles (Polani 1997 p. 119).

El propio Tjio en colaboración con Theodore Puck repasó los resultados, y juntos comprobaron en 1958 que ese número 46 era el usual, no importaba cual fuera el órgano del cuerpo humano cuyos tejidos se estudiaran.

La determinación del sexo

El sexo humano como realidad citológica pertenece a los primeros tiempos de la historia de la genética clínica. La identificación de los cromosomas sexuales humanos se produjo en la década de 1920, aunque no hubo aun acuerdo. Pese a discutirse la posibilidad de que el sexo masculino estudiado en cultivos de biopsias testiculares fuera el par XO, como se asumió en un principio para la mosca *Drosophila*, otros citólogos propusieron el par XY, la discusión de cuya identificación pudo deberse al pequeño tamaño del cromosoma Y (Harper, 2006, pp. 9-10). Theophilus S. Painter, que había observado 48 cromosomas en los dos sexos, fue uno de los primeros en atribuir al par XY el sexo masculino humano (Delgado Echeverría 2007: 582).

En 1949, Murray Barr y su estudiante de doctorado Ewart Bertram identificaron en la Universidad de Western Ontario en Canadá un cuerpo que se teñía intensamente en el núcleo de las células de gato. Aquel cuerpo extraño aparecía en hembras y estaba ausente en machos de gato (Miller 2006, Barr y Bertram 1949; Potter y Soltan 1997). Barr y otro de sus colaboradores, Keith Moore, mostraron que también en tejidos humanos era posible identificar mujeres y hombres en función de la presencia o ausencia de lo que comenzó a llamarse también cromatina o cuerpo de Barr. El propio Moore desarrolló un método que permitía usar muestras procedentes del raspado

con una espátula dental del interior de la boca, citología bucal en la que podía identificarse la cromatina sexual (Moore y Barr, 1955; Ha 2011).

La identificación citológica del cuerpo de Barr permitió determinar el sexo citológico, lo cual contribuyó a resolver casos dudosos, como el de mujeres que no tenían menstruación o que no se embarazaban, y casos de intersexualidad (hermafroditas entre estos). El cuerpo de Barr se convirtió en privilegio epistémico al aceptarse como árbitro del "verdadero" sexo en cada caso (Miller 2006).

La citología del sexo humano como realidad citológica se convirtió de inmediato en asunto de interés para clínica. Paul Polani, en la Unidad de Investigación Pediátrica del Hospital Guy's de Londres, había comenzado a estudiar las enfermedades cardiovasculares congénitas en mujeres con el síndrome de Turner. Polani creía que estás eran genéticamente varones, puesto que las enfermedades del corazón se manifestaban en hombres con mucha mayor frecuencia que en mujeres y sus pacientes con ese síndrome tenían las gónadas muy pequeñas. La formulación del asunto muestra fronteras entre ontologías que estaban siendo atravesadas desde los orígenes mismos de la medicalización de la determinación sexual. El síndrome de Turner se diagnosticaba en mujeres de baja estatura o enanas que carecían de gónadas o las tenían de tamaño muy pequeño, carecían de menstruación y tenían unos caracteres sexuales secundarios característicos, lo cual no afectaba a su inteligencia. Polani también estudió el síndrome de Klinefelter , que se diagnosticaba en varones de apariencia normal pero infértiles (Lennox 1961).

Polani envió al laboratorio de Charles Ford en Harwell muestras de médula ósea de pacientes con ambos síndromes, que tenían en común problemas de infertilidad. Ford, en colaboración con Patricia Jacobs determinó casos de Klinefelter, uno ellos varón con cromosomas XX –pareja ya entonces asignada al sexo mujer-. Encontraron también diagnósticos citológicos con dos cromosomas X y uno Y (XXY) para el mismo síndrome. En 1959, Ford, Polani y un grupo de colaboradores de ambos encontraron la anomalía cromosómica XO es decir, un solo cromosoma X, en una mujer diagnosticada como Turner (Harper 2006). Las muestras viajaban provistas de diagnóstico clínico entre los laboratorio de Polani y Ford, y en esos tránsitos se definía y redefinía el sexo, y su correlación con la anatomía de la persona de la que procedían las muestras. El laboratorio y las prácticas citológicas con extractos, en busca de células que pudieran verse al microscopio y de métodos que rompieran la membrana celular y liberaran los cromosomas, reconstruyó el diagnóstico de estos síndromes, y participó en un proceso de invisibilización de los cuerpos. Esa invisibilidad acarreaba autonomía para los cromosomas, y concedía superioridad al diagnóstico citológico respecto del clínico: el sexo real eran esas combinaciones que ofrecía el cariotipo, el conjunto de cromosomas identificados.

La circulación de muestras, diagnósticos y métodos de hacer cariotipos atravesaba las fronteras entre laboratorios y problemas científicos y médicos –problemas que eran biológicos para Ford y Hamerton en Harwell, eran médicos para Polani en el Guy's Hospital. Esa circulación contribuyó a resolver un rompecabezas en el que piezas biológicas encajaban con piezas médicas. Lo cual proporcionó a Polani confirmación de sus hipótesis sobre el Turner como un síndrome XO, que explicaría la falta de ovarios en las pacientes y contribuía a la construcción de un acuerdo entre la evidencia disponible sobre el carácter masculino de las enfermedades del corazón y la falta de uno de los dos cromosomas X propios de las mujeres. La asignación de valores de género a esta práctica y al estilo de razonamiento permite identificar este acuerdo.

El syndrome de Down

El "mongolismo" fue uno de los primeros desórdenes cuya identificación cromosómica se produjo de manera temprana, simultánea a la de los desórdenes que afectaban a los cromosomas sexuales. Aunque en 1960 un grupo de investigadores y médicos solicitó en una carta a la revista médica británica The Lancet, un cambio en la denominación del desorden, para aliviar su carga racista, y recomendar el de Down o síndrome de Down, el termino mongólico, o mongoloide siguió usándose durante años en algunos foros.

Había sido J. Langdon H. Down quien en 1866 describió las anatomías y desórdenes que afectaban a un grupo de "idiotas" del Asilo Royal Earlswood en Redhill, Surrey. En él informaba de que "un gran número de idiotas congénitos (de nacimiento) eran típicos mongoles. (…) Se parecen tanto entre sí que cuando se les mira es difícil creer que no son todos hermanos (Langdon Down 1866, Wright 2011). Los caracteres que Down les atribuyó cuando describió sus rasgos sobrevivieron como categorías clasificatorias: pelo oscuro, cara plana, ojos oblicuos, labios largos y finos, lengua larga, nariz pequeña, y sobre todo, unas habilidades mentales menores a quienes eran de su edad considerados "normales". Lionel Penrose, uno de los más influyentes estudiosos de la genética de poblaciones humanas – precitológica -, empezó a estudiar este desorden en niños del asilo de Colchester (Royal Eastern Counties Institution) a partir de 1931, cuando fue designado para realizar una investigación sobre la naturaleza y las causas del retraso mental. El retraso mental era, sobre todo, según Penrose, una categoría administrativa para admitir en instituciones de asilo a quienes así se clasificara.

Cuando se instaló en el Laboratorio Galton, Penrose lo convirtió en un observatorio de este tipo de desórdenes asociados a retraso mental. La biopsia seguía siendo la técnica con la que extraer muestras para describir cromosomas de una persona. Y ese fue el tipo de muestra en el que Marthe

Gauthier puso a punto, a petición de Jerome Lejeune y Raymond Turpin, una técnica de cultivo de tejidos que le permitió identificar a los nueve niños "mongólicos" como poseedores un cromosoma adicional al considerado normal. Con 47 cromosomas aquellos cinco niños y cuatro niñas del Hospital des Enfant Malades de París proporcionaron un numero de cromosomas para el síndrome de Down.

Las líneas de la palma de la mano y los rasgos faciales habían hecho de la observación anatómica el origen del diagnóstico de este síndrome. Los patrones dermatoglíficos, de líneas de la mano y del pié cuyas huellas se conservaban pintando las palmas de manos y pies de quien lo padecía, quedaban impresos en el papel para la historia clínica. Aquellas formas faciales y huellas digitales se encontraron con una lesión en el núcleo de las células somáticas: un exceso de cromosomas generaba un defecto congénito.

Tras conocer los resultados del grupo francés, Polani y Ford también hicieron cariotipos a niños y niñas con síndrome de Down y describieron algunos casos infrecuentes de herencia de este síndrome que Daniel Kevles (1987) ha considerado como uno de los orígenes de la genética clínica.

El líquido amniótico

Desde mediados de la década de 1950 comenzaron a difundirse métodos para analizar las células fetales suspendidas en el líquido amniótico extraído del abdomen de mujeres embarazadas. La muestra se extraía –como hoy, por punción en el abdomen de la mujer embarazada, del que se extraían entre 10 y 20 cc. Aunque este método se habñai usadoi previamente para eliminar exceso de líquyido, su finalidad diagnóstica está asociada al término con el que se le conoce desde la década de 1960: amniocentesis. Los primeros diagnósticos lo fueron del sexo del feto, pero muy pronto el riesgo de retraso mental en casos de síndrome de Down fue uno de los diagósticos más frecuentes en mujeres de más de 35 años, con historias previas de abortos espontáneos repetidos en su arbol familiar (Santesmases 2008).

En 1968 se describió el análisis de células fetales de síndrome de Down a una muestra amniótica y se informaba de que, una vez se hizo esta identificación, se interrumpió el embarazo. El análisis del líquido amniótico permaneció, desde que se inició esta práctica a mujeres embarazadas, asociado a la legalización del aborto (Schwartz-Cowan 2001). Uno de los pioneros en ella, Henry Nadler, del Chicago Children's Memorial Hospital, consideró "de gran valor" que se permitiera el aborto en casos de síndromes identificadas en células fetales. El sexo fetal pudo ser identificado en esas muestras, de forma "infalible" haciendo el cariotipo fetal. En 1969 para algunos genetistas clínicos aun pioneros la única indicación para realizar estudios de células fetales por amniocentesis

era la historia reproductiva de la madre, pues la incidencia del sindrome de Down aumentaba "muchísimo" en madres que se acercan a los 40 años.

Fue en 1975 cuando quedó aceptado en Estados Unidos que la práctica de la amniocentesis podía conducir a la interrupción voluntaria del embarazo si se habían detectado anomalías cromosómicas asociadas a enfermedades o a retraso mental (Culliton 1975)

Desórdenes cromosómicos: nuevas lesiones con privilegio epistémico

Pese a que la diversidad de combinaciones de cromosomas XY se resistieron a una correlación simple con la dicotomía hombre-mujer, tanto como a una asignación sencilla y estable del par XX a las mujeres y el XY a los hombres, esa diversidad no arrrojó en ese momento todavía dudas sobre si con ella se designaba un sexo verdadero o simplemente una circunstancia citológica, como ocurriría después en muchos otros casos sin relevancia anatómica y clínica.

Unos mililitros de sangre o de médula ósea representan cuerpos. Las técnicas de cultivo les proporcionan independencia respecto a su procedencia. Abandonados a su suerte por ausencia de quien lo cultivara, crecerían o morirían según hubiera sido la capacidad del medio en el que se sembraron. La metáfora agrícola, como ha dicho Landecker, refiere a la producción, que quiebra la individualidad del material en estudio: el proceso de cultivo es multiplicador y en la multiplicación el sistema olvida la procedencia de la cepa original. Esa invisibilidad parece inseparable del proceso de medicalización y tecnificación de los procedimientos de asignación de salud y de enfermedad.

Esas son las bases sobre las que parece asentarse la dualidad mujer/hombre, que buscó en la citología un sustento para su naturalización, asociada a culturas citológicas que contribuían también a la reasignación del retraso mental.

REFERENCIAS

Culliton Barbara J. 1975, 'Amniocentesis: HEW Backs Test for Prenatal Diagnosis of Disease', *Science* 190, 537-540.

Delgado, I. 2007, *El descubrimiento de los cromosomas sexuales. Un hito en la historia de la biología*, Madrid: CSIC.

Gaudillière, J.-P. 2001, 'Bettering Babies: Down's Syndrome, Heredity and Public Health in Post-war France and Britain', en I. Löwy y J. Krige (eds.) (2001) *Images of Disease. Science, Public Policy and Health in Post-war Europe,* Luxemburgo, Office for Official Publications of the European Comunities.

Gauthier, M. 2009, 'Cinquentenaire de la trisomie 21. Retour sr une decouverte', *Medicine/Sciences* 9 25, 31-316.

Ha, N. Q. 2011, The Riddle of Sex: Biological Theories of Sexual Difference in the Early Twentieth-Century, *Journal of the History of Biology* 43, 215-264.

Harper, P. 2006, *First years of human chromosomes. The beginning of human cytogenetics*, Bloxham: Scion.

Harper, P. 2007, 'Paul Polani and the development of medical genetics', *Human genetics*, 120, 723-31.

Kevles, D. J. 1985, *In the name of eugenics. Genetics and the uses of human heredity,* Berkeley-Los Angeles: University of California Press.

Kottler, M. 1974, 'From 48 to 46: cytological technique, preconception and the counting of the human chromosomes', *Bulletin of the History of Medicine* 48, 465-502.

Lennox, B. 1961, 'Chromosomes for beginners', *Lancet* 1, 1046-1051.

Lindee, S. M. 2006, *Moments of truth in genetic medicine*, Baltimore MD: Johns Hopkins University Press.

Martin, A. 2004, 'Can't any body count? Counting as an epistemic theme in the history of human chromosomes', *Social Studies of Science*, 34: 923-948.

Miller, Fiona A. 2006, 'Your true and proper gender': the Barr body as a *good enough* science of sex', *Studies in History and Philosophy of Biological and Biomedical Sciences* 37, 459-483.

Santesmases, M. J. 2007, 'Hacia descendencias saludables: algunos orígenes del diagnóstico prenatal en España', *Asclepio* LX, 129-150.

Santesmases, M.J. 2009, 'Extractos: cuerpos invisibles en la muestra", *El rapto de Europa* 14, 23-28.

Schwartz-Cowan, R. 2001, 'Medicine, Technology and Gender in the History of Prenatal Diagnosis', en A. Creager, E. Lunbeck y L. Schiebinger (eds.), *Feminism in Twentieth-century Science, Technology and Medicine,* Chicago: Chicago University Press.

Wright, D. 2011, Downs:The history of a disability, Oxford: Oxford University Press.

ANA Y MIA. ESPACIOS FLUIDOS, ESPACIOS POLÍTICOS Y CONTROVERSIAS MÉDICAS

MARÍA GONZÁLEZ AGUADO
Universidad de Viena

ABSTRACT: Ana and Mia. Fluid Spaces, Political spaces and Medical controversies
The present text develops an ontological analysis of anorexia and bulimia nervosas with the aims of shedding a light on the controversies regarding their aetiolology as well as making sense of the so-called Ana and Mia groups' narratives; and finally, deconstruct the supposed normality of the virtual communities aimed to loose weight *on-line*. The article will try both to demonstrate that Ana and Mia's knowledge production can be understood as a political claim and explain the reasons why pro-diet groups do not generate either medical or social warnings in spite of their potential danger.
For doing so, he mentioned syndromes will be conceptualized as being hybrid entities able of embodying and generating new meanings across cyberspace applying Haraway's notion of "material-semiotic system". It will be explained how eating disorders on the Internet are the enactment of different objects within different contexts instead of being just controversial objects for medicine and society.
KEY WORDS: eating disorders, Ana/Mia, ontology, material-semiotic actors

Las comunidades virtuales denominadas Ana y Mia reciben su nombre de la personificación femenina de los trastornos de la conducta alimentaria (TCA) más conocidos: la anorexia y la bulimia nerviosas (A/B). Estos grupos *on-line* son definidos comúnmente como un movimiento unitario que hace apología de los mencionados síndromes argumentado que, en realidad, son estilos de vida o una elección individual, lo que supone una homogenización errónea y la caracterización prejuiciosa del "conocimiento experiencial" de dichas comunidades; además de ser la justificación de su ataque y persecución por parte de la experticia biomédica, los medios de comunicación y algunas asociaciones de pacientes. Como ejemplo de reacciones hostiles hacia Ana y Mia, cabe citar la campaña llevada a cabo por la asociación *Protégeles* –originalmente dirigida a combatir la pornografía infantil- en el año 2005. Esta campaña desembocó en la movilización de los usuarios de Internet en contra de esta comunidad a través de un llamamiento público para denunciar sus contenidos digitales que condujo a la clausura de un gran número de sus páginas *web* por parte la policía asegurando que incitaban a desarrollar estos trastornos.

Disciplinas como los Estudios de Enfermería (Nursery Studies), la Antropología Visual y la Psicología del Género, por el contrario, ofrecen una conceptualización del fenómeno Ana y Mia

más exhaustiva y cercana a la realidad que la versión que enarbolan el binomio biomedicina-medios de comunicación. Gracias al análisis llevado a cabo desde saberes críticos con el "paradigma epidemiológico dominante" (Brown *et al*. 2001, 248) de estos síndromes –ejemplificado en las categorías diagnósticas de los TCA recogidas en la cuarta edición del *Manual Diagnóstico y Estadístico de los Trastornos Mentales* (DSM IV-*Tr*)- podemos saber que la proliferación de estas comunidades es una respuesta a la clausura sistemática de sus páginas y que su carácter es nomádico y heterogéneo, hasta el punto de que puede afirmarse que dentro de ellas existen tantas definiciones y valoraciones de los TCA como usuarias de estos grupos: desde la celebración de la A/B hasta su definición como enfermedad y el reconocimiento del daño que provocan (Gilles 2003, 464-468). De la misma manera, estos enfoques no normativos revelan que la finalidad de Ana y Mia es muy distinta a la de la promoción de las enfermedades que nos ocupan. Dias propone, en este sentido, que, si bien su propósito inicial es ofrecer materiales para perder peso y ahondar en la experiencia de estas enfermedades de manera comunitaria, estos contenidos digitales suelen evolucionar hasta convertirse en un espacio de apoyo para las usuarias que no han logrado recuperarse todavía (Dias 2003, 34). Pero la razón última de la emergencia de este colectivo es mucho más grave y compleja: Ana y Mia son un reflejo del complicado estatus epistémico del "paradigma epidemiológico dominante" de los TCA que, al margen de no dar cuenta de la extensa variedad de los casos clínicos ni explicar la paradoja de que una mayor difusión preventiva de estas patologías redunde en un incremento de su prevalencia, se basa, reproduce y legitima estereotipos sexistas, étnicos y coloniales como han demostrado, entre muchas otras autoras, Helen Gremillion, Maree Burns, Susan Bordo y Kim Chernin.

Dias afirma que estas comunidades –altamente feminizadas en el momento de la publicación de su artículo- pueden ser conceptualizadas como un acto de transgresión en contra de los discursos biomédicos y psiquiátricos dominantes acerca de estos síndromes. Dichos discursos describen a las mujeres que los padecen como personas irracionales que persisten en negar su enfermedad y, a la vez, medicalizan sus experiencias. Ana y Mia reproducen y resisten, a la vez que negocian, los discursos culturales que definen sus cuerpos, y los del resto de las mujeres, con el objetivo de subvertir dicha apropiación sexista de sus cuerpos a través de sus narrativas. (Dias 2003, 31).

Resulta a primera vista contradictorio que el conocimiento lego articulado por Ana y Mia conduzca a resultados tan lógica y mutuamente excluyentes como son reproducir y, a la vez rechazar, el ideal de corporal de la delgadez. Pero este aparente sinsentido no es si no el fruto de la simplificación, por parte del feminismo liberal y reformista de corte ilustrado, de la infinidad de maneras en las que la lucha antipatriarcal se materializa. Dicho reduccionismo invisibiliza aquellas acciones

subversivas del orden patriarcal que no sean la oposición "racional" y la negación pública del sistema de sexo/género. Sin embargo, si analizamos las narrativas de estas comunidades aplicando la noción de "esencialismo estratégico" (*strategic existialism*) (Braidotti 1994, 4), dichas narrativas se vuelven coherentes y, lo que es más importante, claramente antipatriarcales. El "esencialismo estratétigo" braidottiano concibe la radicalización y exageración de los imperativos de género como una herramienta ético-política que subvierte el orden patriarcal a través de la visibilización de sus consecuencias más extremas. Por lo tanto, siguiendo a Bradiotti, la performatividad comunitaria de A/B dirigida a la encarnación del ideal corporal femenino occidental llevada a cabo por estos grupos termina siendo, en realidad, una denuncia implícita de su crudeza y poder de sujeción a través de la encarnación de sus consecuencias últimas: una voluntad limitada a la consecución y mantenimiento de un cuerpo emaciado.

Este análisis de la experiencia comunitaria Ana y Mia se desmarca de los discursos biomédicos que minimizan e invisibilizan la influencia de los modelos corporales sexistas en la génesis y aumento de la prevalencia de los TCA y, lo que es peor, reifican la estigmatización de las mujeres que padecen estos síndromes al caracterizar la experiencia de su enfermedad como irracional y desviada. Pero esta interpretación del "conocimiento experiencial" (Borkman 1976, citado por Akrinch 2010, 4.3) de los grupos Ana y Mia en modo alguno puede conducir a la relativización de la peligrosidad de sus prácticas corporales. No se puede obviar que en la producción digital de estas comunidades –en los formatos de foro, página *web*, *blog* y perfil o grupo en las redes sociales- abundan ejemplos de dietas extremas e incluso ayunos, que se recomienda el uso de medicamentos para adelgazar sin prescripción médica y se recomienda el ejercicio físico excesivo, a modo de ejemplo. Ana y Mia, asimismo, ha llevado a cabo una re-interpretación de los sistemas biomédicos de monitorización corporal como son el índice de masa corporal (IMC) y las tablas de peso y altura reduciendo radicalmente los valores que definen el "peso normal" o saludable. Por otro lado, toda reconstrucción narrativa de la enfermedad, como la realizada por Ana y Mia, supone una resistencia a las ideologías normalizadotas, como ha recogido Gray para el caso de los padres de niños con autismo (Gray 2001, 1248). En este sentido, se puede interpretar los contenidos Ana y Mia como una negación del estatus patológico de los TCA y, por tanto, justificar la alarma que generan. Sin embargo, la coherencia y neutralidad de estos discursos dejan de ser tales cuando entran en escena otro tipo de comunidades virtuales también encaminadas a la pérdida de peso en grupo; las cuales preconizan la misma idea de cuerpo como proyecto que Ana y Mía y los mismos métodos para alcanzarla, pero que no han sido objeto de crítica ni de alarma sanitaria o mediática alguna, son los grupos virtuales pro-dieta.

Defino los grupos pro-dieta como comunidades virtuales que trasladan la experiencia del adelgazamiento y la reducción o modificación de la forma corporal de la esfera íntima, al ámbito de lo público a través del uso de herramientas digitales. La existencia de estos grupos responde a la necesidad de compartir y obtener apoyo para concluir de manera exitosa un plan de adelgazamiento generando, de esta manera, "conocimiento experiencial" (Borkman 1976, citado por Akrich 2010, 4.3) acerca de este proceso. Existen dos tipos de comunidades pro-dieta: las que pueden definirse como "comunidades de práctica" (*communities of practice*) y las que responden al modelo de "movimiento orientado hacia un producto o tecnología" (*technology and product-oriented movement*). Las "comunidades de práctica", según las define Akrich, se caracterizan por compartir una definición de las bases que constituyen el grupo y por establecer relaciones grupales sencillas e intensas. La autora añade que este tipo de comunidades desarrollan una serie de objetos o productos como resultado del intercambio de experiencias individuales (Akrich 2012, 2.1-3.6). La diferencia de éstas respecto a los "movimientos orientados hacia un producto o tecnología" es que estos últimos son movilizaciones asociativas en alianza con el sector privado, con el objeto de promover el uso de una determinada tecnología o producto en la que los miembros perciben cierto potencial de cambio social (Hess 2005, 516). Este es el caso de aquellos grupos pro-dieta que preconizan el uso de una determinada dieta como método definitivo para perder peso, dietas que se oponen y contradicen el enfoque biomédico actual consistente en la reducción de la ingesta calórica, la prohibición de ciertos alimentos y la práctica de ejercicio físico. Los grupos pro-dieta con mayor presencia en la Red son actualmente los que promocionan dietas proteicas como la Atkins y la Dukan.

Como ya se ha señalado, las semejanzas entre las comunidades Ana y Mia y las pro-dieta son más que evidentes. No solamente comparten ideales y prácticas corporales, si no que el análisis de la producción narrativa de esta última revela pensamientos y comportamientos que bien podrían ser definidos como síntomas de TCA. Las coincidencias entre ambas llegan al punto de compartir las mismas dietas y productos adelgazantes y, lo que es más grave, la etnografía virtual de grupos Ana y Mia demuestra que las usuarias de estos últimos visitan y se inspiran también en los grupos pro-dieta. Aunque es cierto que los servicios públicos de sanidad han alertado acerca del riesgo de las dietas proteicas citadas, por qué no los grupos pro-dieta no han generado alarma social, ni se ha calificado a sus miembros como personas enfermas o desviadas y por qué no se clausuran sus páginas.

En la controversia en torno a Ana y Mia, los TCA han sido tratados como *boundary objects* (Leigh Star y Grisemer 1989, 393), como objetos lo suficientemente maleables como para que

diferentes comunidades desarrollen diferentes definiciones de los mismos y, a la vez, conserven su forma e identidad. Del mismo modo, en que esta noción resultó inadecuada para el análisis de la cirrosis hepática realizado por Singleton y Law (Singleton y Law 2002), tampoco lo es para el caso que nos ocupa. Quisiera proponer en este texto la conceptualización de estos síndromes como "sistemas semiótico-materiales", noción que Haraway utiliza para definir entidades como el ADN o las células cancerígenas en *Las Promesas de los Monstruos* (Haraway 1992). Los TCA, por lo tanto, son *híbridos* de naturaleza y cultura, con agencia y presencia material. De este modo, se puede explicar la controversia científica sobre la etiología de los TCA que enfrenta al "paradigma epidemiológico dominante", que minimiza la influencia de la cultura en su génesis y prevalencia, con los estudios realizados desde las Ciencias Sociales, que la relacionan con el ideal sexista y colonial de la delgadez.

Sin embargo, los TCA son "sistemas semiótico-materiales" en un sentido más contundente: son susceptibles de encarnar y generar nuevos significados, independizándose así del ámbito biomédico, en otras palabras: los TCA son más que una categoría diagnóstica. Tal y como reflejan las etnografías virtuales de grupos Ana y Mia, sus narrativas muestran la existencia de una "posición anoréxica de sujeto" (*anorectic subject position*) previa a la existencia del grupo, articulada a partir de materiales provenientes de los medios de comunicación y del conocimiento público acerca de estos síndromes. El análisis etnográfico también revela que las personas que participan en dichos colectivos han re-interpretado las categorías diagnósticas del *DSM* como un decálogo para convertirse en la "perfecta anoréxica", es decir, para encarnar rigurosamente esa posición de sujeto (Allen 2008, 591-2). Por lo tanto, el análisis de los TCA debe ser –siguiendo la propuesta de Mol- de carácter *ontológico*: la controversia científica y mediática acerca de Ana y Mia no se debe a que distintos colectivos mantengan conceptualizaciones diferentes de una misma entidad, si no que A/B son la materialización de diferentes objetos en diferentes contextos y haces de relaciones.[2]

Los grupos pro-dieta definen sus prácticas corporales como saludables, identificándose con los discursos biomédicos que combaten la obesidad. En ellos, las referencias a A/B son continuas, pero siempre como estrategia discursiva para justificar la saludabilidad de dichas prácticas, aunque éstas sean potencialmente peligrosas y, a veces, las mismas que preconizan Ana y Mia. Los TCA se transforman en diferentes objetos en el *cyberespacio*, entre otros, pueden ser identidades grupales, en el caso de las comunidades Ana y Mia, o entidades normalizadoras de los estereotipos corporales

[2] La definición usada del "análisis ontológico" de AnnMarie Mol está recogida en el artículo de Vicky Singleton y John Law "Object Lessons", en su versión de 2 de Julio de 2004 (Singleton & Law 2004, 10-11, accesible en: http://www.heterogeneities.net/publications/LawSingleton2004ObjectLessons.pdf

occidentales y del discurso biomédico que combate la obesidad en los grupos pro-dieta. De ahí, que no generen controversia ni pánico. Y esta capacidad de A/B para transformarse en diferentes objetos es posible gracias a que Internet es un "espacio fluido" (*fluid space*), que permite la emergencia y transformación de redes y comunidades y la materialización de objetos múltiples; al igual que el espacio en el que "fluía" la anemia en el análisis desarrollado por Mol y Law (Law y Mol 1994, 658-664).

REFERENCIAS:

Akrich, M. 2012, "From Comunities of Practice to Epistemic Communities: Health Mobilizations on the Internet", *Sociological Research Online*, 15(2), 10

Allen, J, 2008, "The Specularization of the Anorexic Subject Position", *Current Sociology*, 5, 587-603

Braidotti R. 1994, *Nomadic Subjects: Embodiment and Sexual Difference in Comtemporaty Feminist Theory*, New York, Columbia University Press

Brown Ph. *et al.* 2001, "A Gulf of Difference: Disputes over Gulf War- Related Ilnesses", *Journal of Health and Social Behaviour*, 42(3), 235-257

Dias, K., 2003, "The Ana Sanctuary: Women's Pro-Anorexia Narratives in Cyberspace", *International Journal of Women's Studies*, 4(2), 31-45

Gilles, D. 2006, "Constructing Identities in Cyberspace: The Case of Eating Disorders", *British Journal of Social Psychology*, 45, 463-477

Haraway, D. 1992, "The Promises of Monsters: A Regenerative Politics for Inappropriate/d Others", en: Lawrence Grossberg, Cary Nelson, Paula A. Treichler, eds., *Cultural Studies*, New York, Routledge, 295-337

Hess, D. J. 2005, "Technology-and Product-Oriented Movements: Approximating social Moement Studies andScience and Technology Studies", *Science, Technology, & Human Values*, 30(4), 515- 535

Law J. y Mol A. 1994, "Regions, Networks, and Fluids: Anaemia and Social Topology", *Social Studies of Science*, Nov. 1994, 641-671

Law J. y Singleton V. 2004, "Object Lessons", version de 2 de Julio de 2004, accessible en: http://www.heterogeneities.net/publications/LawSingleton2004ObjectLessons.pdf

Leigh S. y Griesemer J. R. 1989, "Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39, *Social Studies of Science,*19(3), 387-420

CUERPOS EN TRÁNSITO: BIOPOLÍTICAS DE LA CIENCIA Y LA TECNOLOGÍA II

Coordinadoras:
MARÍA JESÚS SANTESMASES y EULALIA PÉREZ SEDEÑO
Instituto de Filosofía-CCHS-CSIC

Este simposio es continuación del tiulado "Cartografías del cuerpo: Biopolíticas de la ciencia y la tecnología". Como él, presenta los resultados de un proyecto de investigación, financiado por el Plan Nacional I+D+I, FFI 2009-07138-FISO, en el que, durante tres años, se han analizado, desde una perspectiva de Ciencia, Tecnología y Sociedad, las ciencias y tecnologías del cuerpo, el estatuto que los cuerpos ocupan y desempeñan en las prácticas científicas. A través del estudio de un conjunto de tecnologías (terapéutico-reproductivas, de asignación y reasignación de sexo y de mejora y reforzamiento de los cuerpos) hemos examinaddo las representaciones de los cuerpos sexuados, y en particular los de las mujeres, el papel que desempeñan en la tarea científica, y como receptores y proveedores privilegiados de biomateriales para esas tecnologías, al tiempo que son, resultado de los procesos y relaciones promovidos por éstas.

FENOMENOLOGÍA DESDE LOS CUERPOS PATOLÓGICOS[1]

MARÍA JOSÉ MIRANDA SUÁREZ
USC

ABSTRACT: Phenomenology from pathological bodies
The body (Leib, Chair and Körper) is one of the main issues that will begin to take relevance, particularly in the French phenomenology of 20th century. In particular, the analysis of the tension of the lived body, and its reflection in pathological situations, is fundamental to distinguish phenomenological as Leib and Leibkörper concepts, or the need to rescue the notions of «body scheme» and «body image». We will deepen the influence that Psychiatry and Neurology has had on the development of the phenomenology of the body. We articulate this speech as a place of anchorage not only of feminist criticism, but also of posthumanisms.
KEY WORDS: phantom limb, body integrity identity disorder, posthumanism, phenomenology

1. Leib, Körper, Chair

Pese a que en sus inicios la fenomenología y psiquiatría han compartido contextos de génesis[2], posteriormente la filosofía ha tendido a dejar estas temáticas al ámbito de la psicología y algunas corrientes literarias. En el libro segundo de *Ideas*, el cuerpo es sobre todo el cuerpo vivo (*Leib*): ya no es un cuerpo como un objeto espacial, sino un comienzo, el de la conducta humana y la vida, desde el que las partes del cuerpo están orientadas en el espacio (Husserl, 1997). De este cuerpo que somos (*Leib*), se distingue el cuerpo que tenemos (*Körper*). Sin embargo, "no podemos cortar el cuerpo en dos diciendo 'aquí el pensamiento, la conciencia; allí la materia, el objeto. Hay una profunda circularidad en el cuerpo, a eso yo lo llamo *carne (chair)*'. El cuerpo sólo es un acontecimiento de esa matriz universal, de esa vida en movimiento que porta una reflexividad sensible" (López Sáenz, M. Carmen, 2010: 97). *Körper* y *Leib* están interrelacionados, bien como depósito de nuestros procesos fisiológicos el primero, o como expresión de deseos y proyectos el segundo.

Al relacionarse de manera integral y unitaria *Leib* y *Körper*, apenas se considera su distinción en ocasiones. Frente a este problema, conviene profundizar brevemente en los conceptos de esquema corporal y de imagen corporal presentadas por Jean Lhermitte. Sus trabajos son más

[1] Desde algunas fuentes utilizadas no se crítica el uso de oposiciones rígidas como sano / enfermo, normal / morboso, etc. Algo que no defendemos.
[2] Los autores clásicos fundamentales en los años 50 y 60 son Bingswanger, Minkowski, Boss y May.

neurológicos que fenomenológicos. Sin embargo, no es menos evidente que los materiales y documentos policlínicos, y las conclusiones neuro-fisiológicas que presenta, nos dan acceso a un cuerpo vivo patológico y nos permiten además examinar el estatuto del esquema corporal en la fenomenología. Según Lhermitte, el esquema corporal estará compuesto de imágenes a partir de experiencias desde el nacimiento hasta la muerte: "…se forma como consecuencia de impresiones, sensaciones y de las percepciones sensibles obtenidas en el curso de la vida" (Lhermitte 1939: 23). Conceptualmente, la imagen de nuestro cuerpo debe distinguirse estrictamente de la imagen que puede ser una intuición; ya que, la forma en que la imagen corporal es calculada, según Lhermitte, es inconsciente e imperceptible para la actividad corporal. La imagen emerge a partir de momentos sensibles. La imagen corporal es un componente del esquema corporal y desempeña un papel creativo en el comportamiento. Lhermitte explica a partir del comportamiento de un bebé cómo distingue inconscientemente lo que le parece corresponder con la imagen de su propio cuerpo y lo que no forma parte de dicha imagen. Es un bebé al que le ponemos un guante en la mano derecha, ante lo que comprobamos cómo se sorprende mientras comienza a mover los dedos, dentro del guante, hasta que se lo quita con la otra mano. Son esa serie de gestos, que expresa con su propio cuerpo, los que indican cómo el bebé expresa, por su propio cuerpo, el sentido del guante sin recurrir a la articulación de palabras. El sentido postural brota a partir de la construcción de la imagen del cuerpo, de modo que: "Esta cosa (el guante) no forma parte del bebé, no está dotada de la misma sensibilidad que su cuerpo. Reconocemos aquí el primer esbozo de discriminación entre la imagen del cuerpo y la presentación de un objeto distinto al del cuerpo" (Lhermitte 1939, p. 19). Así, la génesis del sentido postural permite al bebé referirse a dos tipos de campos: el que se refiere a su propio cuerpo y el que le parece «extranjero» a causa de la no pertenencia a su imagen corporal. Y la manera en que el bebé distingue la imagen de su cuerpo y del guante es precisamente un juego: "Si bien a lo largo de nuestras vidas no renunciamos a este instinto lúdico, es indudable que en el niño se realiza más libremente, sin restringir los límites" (Lhermitte 1939, 21 y 22). En este juego, las apariencias y los escorzos del guante no se suceden a unos a otros, sino que son desordenados. En este juego, su actividad es realizada sin propósito y sin necesidad de unificar las apariencias perceptivas y los aspectos espaciales.

Avanzaremos ahora hacia la espacialización de la conducta y la construcción del esquema corporal. Volviendo al ejemplo del bebé, cuando reconoce la imagen de su propio cuerpo, también reconoce partes de su cuerpo. Lhermitte intentó identificar el concepto de esquema corporal desde el punto de vista del desarrollo psíquico. A través de esta facultad de localización, que acabamos de señalar en el bebé, nuestras percepciones pueden integrar un conjunto armónicamente. Lhermitte considera la noción de espacialidad como un rasgo de la construcción del esquema corporal. El niño

comienza a distinguir el sentido postural, su propio cuerpo y el mundo exterior. *A su vez, los datos espaciales son los que también están empezando a caracterizar su acción y comportamiento.* Las sensaciones, hasta el momento dispares y dispersas, comienzan a consolidarse integralmente y a localizarse, generando un espacio propio de comportamiento que funciona como una instancia que atrae la atención del bebé a una cosa u otra. En esta etapa, el sentido postural es espacial. Lhermitte lo denomina sentido del espacio. En esta espacialización, tanto el gesto como la postura refieren al bebé. Donde la construcción del esquema corporal se da en el movimiento por el cual se construye el sentido postural. El esquema corporal tiene la función de favorecer un cierto comportamiento caracterizado por su plasticidad. Es un esquema esencialmente plástico. Según el ejemplo dado, mediante el desarrollo de la acción, la actualización del esquema puede tomar una nueva forma que corresponda a una nueva situación del comportamiento. La construcción del esquema corporal da plasticidad al comportamiento. Permite que el cuerpo del bebé, que seguía siendo una simple parte de la misma organización, acceda a la cultura donde puede entonces libremente corregir su posición y postura.

2. El miembro fantasma

Uno de los casos patológicos que analiza es el fenómeno del miembro fantasma. Es el fenómeno por el cual el cuerpo amputado siente dolorosamente las partes amputadas de sí mismo. A veces es tan fuerte la sensación de los miembros fantasma que parecen más vivos y reales que un miembro sano. El origen de los miembros fantasma es una cierta confusión que el cuerpo amputado conserva con la realidad. Es una experiencia real pero exclusivamente en su conciencia. En algunas personas, el miembro fantasma incluso evitó el contacto con un cuerpo sólido al acercarse al mismo. Lo que indica que todavía hay una cierta espacialidad de los miembros amputados, ya que son capaces de evitar obstáculos. El miembro fantasma puede discernir lo que es sólido de lo que no es, en su movimiento. Eso quiere decir que el paciente conserva una cierta plasticidad en el movimiento cuando encuentra obstáculos. Sin embargo, por paradójico que pueda parecer, son fantasmas y sensaciones, en cierta medida ficticias, los que llenan el espacio. Eso quiere decir que la espacialización del cuerpo amputado es sustituida por la espacialización de los elementos fantasmagóricos.

Desde el punto de vista fenomenológico, las experiencias del cuerpo amputado podrían explicarse a raíz de la distinción entre *Leib* y *Körper*. Para profundizar en ello, es útil recordar que Lhermitte clasifica los movimientos del miembro fantasma en tres tipos: los movimientos espontáneos, las cinestesias y los reflejos. Los movimientos espontáneos son los que se producen sin ser detectado ningún origen inmediato a los mismos. De modo que, el miembro fantasma actúa

de modo incomprensible para el amputado y el observador. Estos movimientos provocan en el amputado ciertas impresiones dolorosas y extrañas. Por su parte, los movimientos cinestésicos van acompañados de movimientos de los miembros sanos. El miembro fantasma, en este caso, refleja un cierto hábito de comportamiento o gesto cuyo origen se remonta a antes de la amputación. Esto significa que la intención del movimiento se mantiene en la vida del cuerpo amputado. Por último, los reflejos son los cambios en la posición de las articulaciones de los miembros fantasma. Los movimientos reflexivos se pueden identificar y fijar en la conciencia del amputado. Este tipo de movimientos, se experimentan sin ningún dolor.

Esta tipología de movimientos, debería darnos algunas claves. Por un lado, identificamos a partir de los movimientos espontáneos, un caso donde la imagen corporal trabaja aisladamente en el comportamiento del amputado. Sus sensaciones dolorosas y extrañas, son proyectadas en dolores del miembro fantasma y, por lo tanto, irreal. Estas experiencias terribles pueden explicarse desde la función creativa de la imagen corporal. La incoherencia se da entre la función creativa de la imagen corporal y la función figurativa del esquema corporal que espacializa un comportamiento más. Esto significa que el aspecto creativo de la imagen corporal se convierte en independiente y autónoma del comportamiento creado. Por lo tanto, es la operación viva pero aislada de la imagen corporal la que vemos en el movimiento espontáneo de la extremidad fantasma. Por otro lado, las cinestesias y los reflejos contrastan con los movimientos espontáneos. Mientras estos indican un aislamiento de la imagen corporal, los primeros revelan un esquema corporal que funciona sin renovarse. Por lo tanto, causan otro tipo de inconsistencia en la articulación entre el comportamiento actual y no renovado del esquema corporal, en cuanto la plasticidad del comportamiento resultante de un esquema no renovado, ya no es capaz de adaptarse al comportamiento actual.

Sin duda, un primer nivel de resistencia del esquema corporal, para mantener la unidad, lo vemos en la construcción de la imagen corporal en los bebés hace necesario que proporcione una estructura, lo suficientemente fuerte como para resistir la dinámica tanto de la perturbación a la que la vida diaria se enfrenta, como los trastornos orgánicos en procesos de enfermedad demasiado frecuentes al principio. Por tanto, el edificio de la imagen corporal es resistente a las perturbaciones. Un segundo nivel de resistencia, se encuentra en el estado patológico del comportamiento, como vimos con el miembro fantasma. Lo que se plantea en última instancia, en este caso, es que nuestra imagen corporal aparece mucho más resistente a la destrucción que nuestra morfología. El cuchillo del cirujano, puede mutilar el cuerpo, quitar parte de su ser físico, pero esa amputación no llega en la misma medida a la estructura de nuestra mente, la cual tarda muchos años en ver lo amputado, en renunciar en el crepúsculo de su subconsciente a la irremisible eliminación de parte de su ser físico, que paradójicamente pretende mantener con todas sus fuerzas.

3. El trastorno de identidad de integridad corporal

En el trastorno de identidad de integridad corporal (BIID) las personas muestran un deseo intenso o bien de ser paralizadas o bien de amputar una o más de sus extremidades sanas. BIID no es una parafilia, ni tampoco un deseo psicótico. Más bien se plantea como un trastorno de identidad. La principal motivación para la modificación corporal se cree que es un desajuste entre el esquema corporal real y el percibido. Actualmente BIID no está incluido en DMS IV. Psicólogos y médicos explican este fenómeno de muy diferentes maneras; pero no se conoce un tratamiento psicoterapéutico o farmacéutico exitoso. Lobbies desde BIID explican el deseo de amputación en analogía al deseo de transexuales para la reasignación del sexo quirúrgico e incluso algunos expertos en ética médicas discuten la controversia acerca de elección de amputaciones de extremidades sanas en relación con las cirugías cosméticas. En 1997, Richard L. Bruno, un especialista en trastornos cerebro-cuerpo en el Hospital de Englewood y Medical Center en Nueva Jersey, publicó una clasificación de apotemnofilias:

1) *wannabes* (aspirantes a amputados) cuyo deseo para amputarse no es erótico sino por la discapacidad en sí

2) *pretenders* que simulan una discapacidad física, por ejemplo, con vendas alrededor de una extremidad y utilizando una silla de ruedas o muletas

3) *devotees* que están sexualmente atraídos por amputados y les gustaría ser así.

Los wannabes y pretenders, argumenta Bruno (1997), buscan reconocimiento y simpatía más que gratificación sexual.

Desde la psiquiatría, el DIIB frecuentemente se da en conjunción con otros trastornos psiquiátricos, la mayoría son diagnósticos comórbidos, trastornos obsesivo compulsivos, de consumo de sustancias, alimentarios y trastornos personalidad. Los tratamientos cosméticos no suelen producir ningún cambio o, peor aún, una exacerbación de los síntomas. Michael First caracteriza el trastorno no tanto como un deseo de discapacidad sino como un trastorno de identidad. En su estudio, casi dos tercios de los sujetos decían que querían una amputación principalmente para establecer su «verdadera identidad» (First, 2004). Compara el BIID con el trastorno de identidad de género GID, en el que las personas se sienten igualmente incómodas con parte de su anatomía. Tales similitudes, sugieren según First que BIID es un trastorno de identidad y deben clasificarse como tales en el DSM, pero recurrir a la definición de BIID como un trastorno de identidad análogo a la identidad de género, no puede considerarse como una clasificación ni descriptiva ni explicativa, ni psicológica ni biológicamente.

Otro de los factores determinantes del DIIB, es la base neurológica. Uno de los principales factores de la tesis de trastorno de identidad, es que supone un trastorno neurológico en las personas

cuando explican el deseo de amputarse como un reflejo de los miembros fantasma. Algunos de los investigadores teorizan que de los resultados de BIID se deriva una distorsión o eliminación de una de las representaciones de la forma general del cuerpo. La imagen corporal es una representación conscientemente accesible de la forma general y estructura del cuerpo. Se deriva de varias fuentes, incluyendo táctiles, visuales y propioceptivas. Las estructuras de la imagen corporal constituyen la base de las creencias sobre uno mismo y, una lesión en un determinado punto en el mapa, podría efectuar un deseo de amputación específica en una ubicación precisa, por ejemplo, encima de la rodilla de la pierna izquierda. Errores periféricos o centrales también pueden perturbar la imagen corporal (Sacks, 1984). En algunos casos, el BIID podría resultar de una lesión periférica. En 1974 el neurólogo Oliver Sacks se hirió gravemente su muslo izquierdo con un toro en las montañas de Noruega. Después de que la herida sanó, no sintió conexión con su muslo y a veces deseaba tener la pierna amputada. Sacks teorizó que tales lesiones podrían en algunas circunstancias interrumpir la comunicación entre la extremidad y el cerebro. Algunos pacientes BIID igualmente recuerdan las heridas de la infancia en la extremidad que poco después se obsesionaron con amputar. Durante la Segunda Guerra Mundial, Leontjew y Zaporozec describieron un síndrome que denominaron "amputación interior", observaron que aproximadamente 200 soldados con heridas y reconstrucciones quirúrgicas de sus manos las percibían como extrañas. Los sistemas sensoriales que han eliminado parte de la imagen corporal, sólo pueden reintegrarla utilizándolas (Leontjew y Zaporec, 1960). Una ayuda adecuada para esto es música, que puede desencadenar movimientos de extremidades que han perdido su inervación (Sacks, 1984). Las intervenciones del esquema corporal, los reflejos neuronales y los mecanismos del miembro fantasma sugieren que la percepción de un miembro fantasma puede ser un fenómeno evolutivamente adaptativa. El dolor fantasma, sin embargo, puede tratar de mantener los constructos globales del cuerpo.

Conclusión

Los miembros fantasma proporcionan información valiosa sobre los mecanismos subyacentes de la conciencia corporal. La complejidad de los fenómenos del miembro fantasma o el esquema corporal explican ciertos fenómenos corporales. Estos sistemas específicos y procesos han recibido poca atención en la investigación, pero actualmente son revisados con una importante orientación futura, sobre todo en prótesis y propiocepción. Como consecuencia, identificamos fenómenos que deshacen una idea tradicional de cuerpo, y favorecen un espacio discursivo material desde el que poder abordar críticamente en el análisis de las biotecnologías. Paradójicamente la fenomenología del cuerpo nos lleva a procesos de descorporeización, contextuales y autorreferentes.

REFERENCIAS

Blom RM, Hennekam RC, Denys D (2012) "Body Integrity Identity Disorder". *PLoS ONE* 7(4): e34702. doi:10.1371/journal.pone.0034702.

Bruno, R. L. (1997), "Devotees, pretenders and wannabes: Two cases of factitious disability disorder". *Journal of Sexuality and Disability,* 15(4): 243–260.

First, M. B. (2004). "Desire for amputation of a limb: Paraphilia, psychosis, or a new type of identity disorder." *Psychological Medicine* 34: 1–10.

Husserl, (1997), *Ideas relativas a una fenomenología pura y una filosofía fenomenológica. Libro segundo: Investigaciones fenomenológicas sobre la constitución*, (traducción de Antonio Zirión Quijano), UNAM, México. [1952]

Leontjew,A. N., and Zaporozec, A.V. (1960). *Rehabilitation and Hand Function*. London, UK: Pergamon.

Lhermitte, J., (1937), "L'image du moi corporal et ses deformations pathologiques", *L'Encéphale*, Paris, G. Doin et Cie Editeurs, vol. 1, pp. 1-24.

López Sáenz, M. Carmen, (2010), "Hermenéutica del cuerpo doliente-dolido desde la fenomenología del sentir", *Investigaciones fenomenológicas: Cuerpo y Alteridad*, UNED, pp. 89-125.

Sabine Müller (2009): "Body Integrity Identity Disorder (BIID)—Is the Amputation of Healthy Limbs Ethically Justified?", *The American Journal of Bioethics*, 9:1, 36-43**.** Sacks, Oliver (1989). *Con una sola pierna*. El Aleph Editores

PRÁCTICAS MÉDICAS DE LA DIFERENCIA SEXUAL: LA REGULACIÓN DE LOS TRÁNSITOS EN CUERPOS *TRANS*

ESTHER ORTEGA ARJONILLA
(USC)

CARMEN ROMERO BACHILLER
(UCM)

ABSTRACT: In this paper we analyse how gender transits in transgender bodies are regulated in order to stabilize sexual difference and make it appear as "natural" and "unchangeable".
In most psychological, biomedical and judicial discourses on transgender identities, as well as in many activists and academic ones, emphasis is primarily settled on a movement from one gender pole to the other. Yet nowadays, there are several activists groups and certain academic tradition which have taken trans identities as models for gender disruption. Those groups criticize this version of sexual difference because it closes the possibility of recognizing those bodies who do not inhabit neither one pole nor the other.
In a time where several voices strongly claim for the despathologization of trans identities, the regulations of gender transits do not seem to disappear, but to rearticulate in a new fashion that perpetuates gender poles as stable nodes of identity. However what counts as pathological seems to be transformed – as it emerges from the proposal of the American Psychiatric Association to modify "Gender Identity Disorder" first to "Gender Incongruence" and then to "Gender Dysphoria" in the forthcoming DSM-5 -.
KEYWORDS: Gender Identity, sexual difference, transgender, transsexuality, medical standards.

1. De la hegemonía del modelo médico de transición como "migración de género"…

Las prácticas médicas en torno a la transexualidad tienen como sustento teórico un modelo de género heteronormativo que se basa en cuatro supuestos: 1) La *monosexualidad,* esto es, cada persona sólo puede tener *un* sexo médicamente asignado; 2) el *dimorfismo de género*, por el que sólo existen *dos* sexos - hombre y mujer - excluyentes entre sí; 3) la identidad de género es *única, constante e irreversible*; por último, 4) existe el requerimiento normativo subyacente de *coherencia entre sexo, género y deseo heterosexual*.

El sistema sexo/género que hemos descrito es el sistema y modelo en que se basan nuestras relaciones sociales, para categorizarnos como seres sexuados y realizar la primera división social, un modelo que permea también las prácticas médicas y que está presente en los procesos de diagnóstico, tratamiento y cuidado desarrollados en torno a la transexualidad. Así, ¿qué es la transexualidad desde una perspectiva médica? Según el DSM-IV se identifica el trastorno de identidad de género como una condición médica caracterizada por "una fuerte y persistente

identificación con el género opuesto, no simplemente un deseo por las percibidas ventajas culturales de pertenecer al otro sexo" (DSM - IV 1994).

1.1. Transexualidad: Breve genealogía de un concepto.

En la literatura médica, en la que hay una gran afición desde finales del siglo XIX por clasificar las "desviaciones" del sexo, el término transexual aparece en los trabajos de los sexólogos de la primera mitad del siglo XX, bien asociado a la homosexualidad Hirsfeld (1923) o en su acepción actual Cauldwell (1940) como señalan diversos autores (Cohen-Kettenis and Gooren 1999; Meyerowitz 2002). David Oliver Cauldwell será el primero en emplear el concepto transexual en la acepción que actualmente ostenta, es decir, para denominar a personas que desean vivir o viven bajo un rol de género que según el esquema binario y heteronormativo de género de nuestras sociedades, es opuesto a su sexo asignado médica y jurídicamente. Esta tradición de los sexólogos de finales del XIX y principios del siglo XX sitúa la cuestión de la transexualidad, medicamente hablando, en el entorno de influencia de la psiquiatría y la psicología. Posteriormente, ya en la década de los años 60 del siglo XX es el endocrinólogo Harry Benjamin quien hace una compilación de casos clínicos e intervenciones médicas sobre los cuerpos en un libro *The Transsexual Phenomenon* (1966) que pasó a ser considerado como el libro de cabecera de los diferentes especialistas médicos y también de las personas transexuales, que encontraron en el libro un manual a seguir para conseguir tratamiento médico. En el tránsito de unos pocos años, los especialistas médicos con conocimiento privilegiado sobre los cuerpos de personas transexuales habían cambiado; ahora eran la endocrinología y la cirugía las que creaban un espacio habitable para estas personas ya que la psicoterapia no era capaz de "curar" el trastorno. No obstante, esta rápida genealogía sobre la producción médica de la transexualidad no puede obviar que las pugnas entre disciplinas por conseguir la hegemonía continúan y que ya en los años 80 la categorización de la *transexualidad* como un trastorno mental aparece en el manual de diagnóstico psiquiátrico más extendido, el DSM-III de la Asociación Americana de Psiquiatría. En posteriores versiones de este manual, DSM-IV, la categoría es sustituida por la de *trastornos de la identidad sexual* (en su versión en español).

¿Por qué analizamos la regulación de los tránsitos de género?

En esta comunicación atendemos a la regulación de estos tránsitos de género ya que en el Estado español, tanto el diagnóstico como el tratamiento de estos tránsitos de género son requisitos necesarios para conseguir el acceso al cambio legal de nombre y sexo (según Ley 3/2007, de 15 de marzo).

1.2. Teorías médico/científicas sobre el origen de la transexualidad

Las teorías médicas y científicas que han buscado la etiología de la transexualidad se pueden dividir en dos grandes corrientes: 1) Por una lado, la que defiende la sexuación del cerebro y el dimorfismo sexual neuroanatómico del mismo, es decir, que tenemos un cerebro diferente dependiendo de si somos hombres o mujeres. Siguiendo esta teoría, la transexualidad se explicaría como una condición neurológica intersexual en la que una o más áreas sexualmente dimórficas del cerebro son inconsistentes con el sexo somático (Zhou, 1995;). 2) Por otro lado, las teorías de corte psicológico/psiquiátrico que defienden que la transexualidad es una especie de trastorno mental, que paradójicamente, no se puede tratar con psicoterapia sino incidiendo sobre el cuerpo, *haciendo* el cuerpo (Mol y Law 2012), transformando el cuerpo para que aparezca como un espacio congruente con el sistema de sexo/género/deseo heterosexual (Butler 2006), para que aparezca como el cuerpo que *somos* y sobre el que tenemos auto-conciencia (Mol y Law 2012). Esta es la línea en la que se apoyan tanto el DSM en sus diferentes versiones como la *Clasificación Internacional de Enfermedades* (CIE). Ambas corrientes tienen defensores y detractores, si bien a día de hoy la medicina no puede, en base a evidencias científicas, dar una explicación contrastada del fenómeno.

En el contexto del Estado español, la clasificación y tipificación de la transexualidad como categoría diagnóstica ha cobrando fuerza en la medida en que, como explicábamos antes, se ha erigido en requerimiento necesario para poder acceder al cambio registral de nombre y sexo. En este mismo sentido, el tratamiento médico de la transexualidad se convierte también un requisito para la modificación legal del sexo ya que la ley exige que la persona acredite: "Que ha sido tratada médicamente durante al menos dos años para acomodar sus características físicas a las correspondientes al sexo reclamado" (Ley 3/2007).

El debate sobre la forma de inclusión de la transexualidad y, por extensión, de lo que la APA considera como trastornos sexuales y de género en la nueva versión del DSM -V (prevista para 2012 y retrasada, en principio, hasta 2013) nos introduce de lleno en una controversia en la que aparecen nuevos actores y en la que se cuestiona algo más que las categorías a nivel discursivo.

En efecto, el debate en el seno de la APA acerca de la sustitución de la categoría *trastorno de identidad de género/ de la identidad sexual* por una nueva categoría denominada *incongruencia de género* en una primera versión y por *Disforia de género* en una segunda revisión excede los marcos del DSM. El cambio de categorías con pocos meses de diferencia remite a una fuerte crítica al nuevo concepto propuesto de *incongruencia de género* ya que según la definición que se hace de éste, la categoría diagnóstica lejos de desaparecer se ampliaba y bajo su influencia se podrían medicalizar todas las posiciones de género no normativas que no encajasen perfectamente en los

polos de género tradicionales, ya que identificaba como "trastornos" todo lo que no se correspondía con esos polos. Esta patologización ha sido ampliamente contestada desde diferentes foros tanto médicos como activistas, lo que ha desembocado en la sustitución de este concepto por el de *disforia de género* en una segunda revisión. En este caso, se incide en la angustia como elemento clave para definir la categoría diagnostica, pero dado que insisten en identificar las posiciones de género incongruentes como potencialmente productoras de angustia –sin preguntarse si dicha angustia es producto de la transfobia social-, la sombra de la patologización continua desplegada sobre las posiciones de género no normativas.

Así, en este modelo médico de transición de género los sexos aparecen como *cajas negras* en las que las posiciones sociales hegemónicas y estereotípicas aparecen naturalizadas, intocables e incontaminadas, como algo previo a las infecciones de lo social. Todo ello se articula para garantizar la estaticidad de la diferencia sexual y, finalmente, la estaticidad del sistema de sexo/género/deseo.

2. …a los nuevos actores: la demanda de despatologización por lxs activistas trans.

En este contexto de cambios en las categorías diagnósticas pero también en los conceptos legales y realidades sociales en torno a la transexualidad, tanto en el ámbito del Estado español (Platero 2011) como a nivel global se hacen oír cada vez con más fuerza los discursos de los colectivos activistas (Missé y Coll-Planas 2010) y de ciertas tradiciones académicas que problematizan el género y, en cierto modo, nos sitúan ante un nuevo paradigma. La apelación que hace Butler (2010) para intervenir críticamente en el proceso de revisión del DSM implica que, en efecto, nuevos actores han pasado a formar parte del proceso, aunque esté por ver aún la capacidad e influencia de estos actores.

En cualquier caso, el proceso de discusión de este DSM -5 ha trascendido ya el ámbito médico y los actores se reconfiguran – aparecen "nuevos" actores (activistas) que se suman a los "viejos" (médicos)-. Los activistas, antes considerados (simples) pacientes se han convertido en "expertos de los expertos" (Epstein 1996). La diferencia con controversias diagnósticas previas es evidente, ya que si antes eran los "expertos", en forma de psiquiatras, endocrinos y cirujanos los que establecían el diagnóstico y el tratamiento en la actualidad encontramos una multiplicidad de nuevos actores emergentes que intervienen en la controversia.

Desde el año 2007 y con apoyo creciente especialmente a partir de 2009 hemos asistido a la aparición de la Red Estatal por la Despatologización Trans, que ha trascendido las fronteras del estado y se ha convertido en una red internacional que lanzó la campaña "Stop Trans Pathologization 2012". La Red no es el único nuevo actor que aparece como experto en esta

controversia, pero nos interesa especialmente ya que pensamos desplaza el debate y lo sitúa en un paradigma diferente.

La Red establece cinco puntos como objetivos principales de sus reivindicaciones:

- "La retirada del TIG de los manuales internacionales de diagnóstico (sus próximas versiones DSM-V y CIE-11).
- La abolición de los tratamientos de normalización binaria a personas intersex.
- El libre acceso a los tratamientos hormonales y a las cirugías (sin tutela psiquiátrica).
- La cobertura sanitaria pública del proceso de reasignación de género.
- La lucha contra la transfobia: el trabajo para la formación educativa y la inserción social y laboral de las persones trans."

En este sentido no solo cuestionarán los criterios y prácticas de diagnóstico y tratamiento sino que exigirán la retirada de los llamados *trastornos de identidad de género* (TIG) de los manuales internacionales de diagnóstico (DSM y CIE) ya que no consideran lo *trans* como un trastorno o una enfermedad, ni orgánica (física), ni mental.

Los objetivos de la Red plantean el debate en unos términos muy alejados de los realizados hasta ahora, ya que incluye a las personas intersex entendiendo las prácticas médicas tanto en personas trans como intersex como reguladoras del orden binario de género. De esta manera amplían los colectivos conectando luchas y ampliando alianzas. Aquí la Red de Despatologización realiza un movimiento interesante ya que realiza el recorrido inverso al tradicional, al sacar fuera del cuerpo el "problema" *trans*. El problema, según la Red, no sería el de una serie de personas que no se ajustan a categorías normativas y cerradas, sino de un sistema que no puede dar cuenta de la variabilidad y la diferencia y necesita de regulaciones médicas para ajustar a los cuerpos a estas categorías.

Además de esto, la Red de despatologización realiza una apuesta por una transformación de las formas de cuidado, planteando alternativas al *Standards of Care* (SOC) que es la pauta clínica en el tratamiento de lo *trans* y hasta este momento único documento utilizado a nivel internacional (González-Polledo 2010). Así, la Red publicó en 2010 un documento alternativo al SOC, titulado *Guía de buenas prácticas para la atención sanitaria a persoas trans en el marco del sistema nacional de salud*. En la guía se establece un nuevo itinerario y unas nuevas formas de cuidado para la intervención médica con personas *trans* en la que se observa un cambio hacia una medicina más orientada al paciente. La lógica del cuidado por parte de la medicina y la lógica de la elección por parte de los pacientes (Mol 2008) se rearticulan en este caso en formas no esperadas. La lógica de la elección se vincula en este caso a los pacientes que con un alto grado de expertización (Epstein 1996) toman decisiones como pacientes activos. Pero es, precisamente mediante su reconocimiento

como pacientes activos capaces de decidir que articulan un nuevo modelo de cuidado, atravesado por una rearticulación de las posiciones, que sitúe en el centro a la persona trans, habilitando espacios de reconocimiento y aprendizaje en los que la comunidad trans se convierte en un eje fundamental. La transexualidad pasa de enfocarse como una patología individualizada a una forma de conflicto social que desborda los limitados y rígidos marcos normativos de género y que por tanto, no requiere sólo de intervención médica, sino también de apoyo y acompañamiento social. Por otro lado, la insistencia en la autonomía y capacidad de decisión de la persona trans demandante de atención socio-sanitaria revoca la normativización y linealidad impuesta por el actual modelo de cuidado y rompe con la necesidad –cuasi-obligatoria- de moverse de un polo de género al otro, sin espacios intermedios.

3. Conclusiones: ¿de un modelo de migración a uno de diáspora?

Así pues, nos encontramos con un cambio de modelo en el abordaje de la cuestión trans que pasaría de lo que algunos autores han denominado un "modelo de migración" a lo que podíamos identificar como un "modelo de diáspora". En el modelo de migración se tendían a reforzar los cuatro supuestos normativos del género a los que nos referíamos al principio, manteniendo los polos de género intactos e identificando la resolución de la transexualidad como un salto de un polo al otro, donde los hábitos sociales, las proyecciones, presentaciones y actuaciones del cuerpo habían de reforzar –a veces de una forma exacerbada- los modelos vigentes. Sin embargo, en el "modelo de diáspora", como identificamos el conjunto de prácticas que emergen en las nuevas controversias en cuanto a formas de diagnóstico y tratamiento, nos encontraríamos ante un cuestionamiento de dichos marcos normativos y de la linealidad y resolución univocas de las posiciones trans. El concepto de diáspora viene del griego "dispersión" y si bien, en sus primeras acepciones se presentaba como una dispersión desde un origen perdido al que se anhelaba volver, en las revisiones críticas del concepto se insiste en que el "origen perdido" aparece como una idealización mítica. Al cuestionar los "orígenes" como míticamente inscritos es posible identificar una profusión de identificaciones y posicionamientos de género que, en todo caso, resultarán tan artificiales y socialmente mediados como los polos normativos de género identificados como naturales. Esto abre la vía a la autodeterminación de la posición de género como un derecho social básico, según recogen los Principios de Yogyakarta (2007).

Referencias:

A.P.A. (Eds.) 2000, *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition, Text Revision, Washington DC, American Psychiatric Association.

A.P.A. 2011, "DSM-5 Development: Gender Dysphoria" URL (Acceso 9/09/2011): http://www.dsm5.org/ProposedRevision/Pages/GenderDysphoria.aspx
Benjamin H. 1966, *The Transsexual Phenomenon.* URL (Acceso 07/09/2005): http://www.wpath.org/journal/www.iiav.nl/ezines/web/IJT/97-03/numbers/symposion/index-2.htm
Butler, J. 2006, *Deshacer el género,* Barcelona, Paidós.
Butler, J. 2010, "Prólogo: Transexualidad, transformaciones", en Missé, M. y Coll-Planas, G. (Eds.), *El género desordenado: críticas en torno a la patologización de la transexualidad*,, Barcelona, Egales.
Cohen-Kettenis P.T. and Gooren L.G. 1999, "Transexualism: A Review of Etiology, Diagnosis and Treatment", *Journal of Psychosomatic Research*, 46 (4), 315-333.
Epstein, S. 1996, *Impure Science: AIDS, Activism and the Politics of Knowledge*, Berkeley & Los Angeles, University of California Press.
España. LEY 3/2007, de 15 de marzo, reguladora de la rectificación registral de la mención relativa al sexo de las personas, *Boletín Oficial del Estado,* 16 de marzo de 2007, 65, 11251- 11253.
González-Polledo, EJ. 2010, "'No sé qué otra cosa podría ser': medicina entre la elección y el cuidado en la transición FTM", en Missé, M. y Coll-Planas, G. (Eds.), *El género desordenado: críticas en torno a la patologización de la transexualidad*, Barcelona, Egales.
Meyerowitz J. 2004, *How Sex Changed: a History of Transsexuality in the United States*, Cambridge (Mass.), Harvard University Press.
Missé, M. y Coll-Planas, G. (Eds.), 2010, *El género desordenado: críticas en torno a la patologización de la transexualidad*, Barcelona, Egales.
Mol, A. 2008, *The Logic of Care: Health and the Problem of Patient Choice*, London & N.Y., Routledge.
Mol, A. y Law, J. 2012, "Acción encarnada, cuerpos actuados: el ejemplo de la hipoglucemia", en Pérez Sedeño, E. e Ibáñez Martín, R. (Eds.), *Cuerpos y diferencias*, Madrid, Plaza y Valdés.
Platero R. 2011, "The narratives of transgender rights mobilization in Spain", *Sexualities,* 14 (5), 597-614.
Red por la Despatologización de las Identidades Trans del Estado español (Eds.) 2010, *Guía de buenas prácticas para la atención sanitaria a personas trans en el marco del sistema nacional de salud*. URL (Acceso 9/09/2011): http://stp2012.wordpress.com
Zhou, N.J. et *al*. 1995, "A sex difference in the human brain and its relation to transsexuality. *Nature* 1995; 378: 6870-5.

APROXIMACIONES A LA "INTERSEXUALIDAD MÚLTIPLE" DESDE UNA PERSPECTIVA CTS Y CTG

SILVIA GARCÍA DAUDER
Universidad Rey Juan Carlos

ABSTRACT: In this presentation we will summarize the main ideas and learning results from one research line of the project "Cartographies of the body": the intersexuality and the biomedical construction of gendered bodies based on the sex/gender dualism. From a CTS and CTG perspective, we will approach the "intersex multiple", not as *a priori*, but as the effect of different "intersex" produced and performed by different practices and contexts and partial connections between them: in the analysis of discourses present in textbooks, journal articles and medical protocols; in the analysis of biomedical technologies and practices of "normalization"; in the socio-scientific controversy about the name change (intersex conditions versus disorders of sex development); in the representation of intersexuality in the media; in the regulation of intersexuality in the sports; in the articulation of intersexuality from different groups or biosocialities related to it (support groups, activist groups, etc..); and in the knowledge and communities management through new communication technologies (Internet, social networks, etc..).
KEY WORDS: Science, Technology and Society; Sex / Gender; Intersex / DSD

Introducción

El objetivo de esta comunicación es recoger las reflexiones y aprendizajes producto de una línea de investigación presente a lo largo de tres proyectos de investigación[3]: la *intersexualidad como fenómeno socio-médico*. Para ello, se hará un recorrido por las diferentes aproximaciones y versiones de la intersexualidad analizadas en los diferentes proyectos. Con ello pretendemos cartografiar una "intersexualidad múltiple" con diferentes implicaciones teórico-epistemológicas y sociales encuadradas dentro de los estudios de Ciencia, Tecnología, Sociedad y Género. Finalmente se propondrán futuras líneas de análisis o el seguimiento de otras en proceso de cambio.

El sexo múltiple y la *generización* del dualismo sexual

El objetivo general de nuestra primera aproximación a la intersexualidad era mostrar *cómo se representan los cuerpos sexuados y qué papel desempeñan en las prácticas científicas y tecnológicas*. En articulación con otras tecnologías socio-médicas como las tecnologías reproductivas, las tecnologías terapéuticas, de diagnóstico prenatal, etc. nos interesaba la representación y conformación de los cuerpos de las mujeres en especial, junto con los presupuestos

[3] "Interacciones Ciencia Tecnología y Sociedad en ciencias biosociales y tecnologías médicas" (2004-2007);"Ciencia y tecnologías del cuerpo desde una perspectiva CTS" (2007-2010) y "Cartografías del cuerpo: Biopolíticas de la ciencia y la tecnología"(MICINN,FFI2009-07138-Fiso).

y valores de género presentes en estas prácticas y tecnologías. Lo interesante, para el caso de la intersexualidad, es que *una tecnología bio-médica se pone al servicio del diagnóstico del sexo/género y de su posterior intervención en función de unos presupuestos sociales de normalidad corporal*: biología, tecnología y sociedad forman un ensamblaje privilegiado para los estudios CTS.

A partir de la revisión bibliográfica de *manuales* donde aparecían las "condiciones intersexuales" como objeto de estudio socializador de futuros profesionales médicos; a partir de *revistas especializadas y actas de congresos* donde se recogen los últimos avances en tecnologías de forma pormenorizada; y, finalmente, a partir del análisis de *protocolos y prácticas* existentes en el contexto español, pudimos analizar la construcción del "sexo múltiple". No sólo eso, los presupuestos socio-culturales presentes en los criterios para determinar la asignación de un sexo (dentro de la combinación de los diferentes componentes) y la intervención posterior para normalizar un cuerpo sexuado acorde. De este análisis concluimos que, *bajo la apariencia de unidad, eso a lo que llamamos sexo natural se compone de múltiples componentes y que no siempre responden a una dualidad* (varón-mujer). En toda esta bibliografía, aparecen fragmentados y representados mediante diferentes tecnologías de visualización bio-médica los diferentes componentes del sexo (cromosómico, hormonal, gonadal, anatómico, en ocasiones genético y neurológico). Bajo el discurso de la "anomalía", la "ambigüedad", "lo impropiamente inacabado", etc. se exponen diferentes casos donde estos diferentes elementos no se alinean de forma clara bajo un único sexo o no responden a patrones dualistas.

Lo interesante, a partir de aquí, es analizar el argumento circular que parte de un dualismo sexual natural y que, cuando la "propia naturaleza" no responde a él, traduce la variabilidad sexual en ambigüedad o alteraciones del desarrollo que deben ser corregidas. Lo que nos propusimos analizar fue: bajo qué presupuestos de cuerpos estándar de varón o de mujer se produce la asignación de un sexo u otro; qué criterios son los determinantes (y cómo se introduce de nuevo la mirada de género socio-cultural en ello); el papel de las propias tecnologías bio-médicas en la elección de dichos criterios y en la posterior intervención; cómo se justifica una intervención normalizadora posterior que ajuste lo máximo posible los cuerpos al sexo asignado (en términos de "necesidad médica"), etc. Desde esta primera aproximación, *pudimos problematizar los dualismos de sexo/género/deseo y la propia distinción entre un sexo biológico y un género cultural*. Es más, pudimos observar no sólo cómo la mirada de género delinea y fija dos sexos, sino que, para hacerlo, presupone que *es menos costoso psicológicamente cambiar los cuerpos sexuados que las normas sociales de género* (García-Dauder, Romero y Ortega, 2007).

Finalmente, y respecto a cuestiones éticas, señalar cómo los propios cuerpos y vidas de las personas intersexuales han sido tradicionalmente *cosificados como recursos donantes de interés*

científico para dirimir controversias naturaleza/cultura, biología/sociedad, herencia/ambiente, etc. Descuidando la experiencia vivida de la persona intersexual, y de su estigmatización a partir del escrutinio médico, lo que ha interesado a la comunidad científica han sido los "casos de estudio" que por su excepcionalidad se convertían en "claves" para defender una postura u otra dentro de los debates teóricos.

De controversias, biosocialidades y reconfiguración del conocimiento experto

En el segundo proyecto de investigación nos interesamos por las diferentes voces de aquellas personas o grupos que han experimentado la intersexualidad en sus vidas desde diferentes vivencias (no sólo médicas). Y ello porque *la gestión médica de la intersexualidad se encuentra en un momento de "crisis de paradigma",* gracias a las demandas y peticiones de cambio realizadas por las propias personas intersexuales. En este caso, la aparición de una tecnología de la comunicación como Internet ha posibilitado no solo el intercambio de información sobre tratamientos, sino la propia colectivización de las diferentes experiencias intersexuales hasta entonces vividas individualmente y en silencio (Chase, 1998). Con ello, la red ha posibilitado una plataforma de generación e intercambio de conocimiento más democrático (más allá de producciones científicas solo asequibles a la comunidad médica) y de reivindicación de cambios en los protocolos. Como hemos señalado en otro lugar, siguiendo a Latour, los "objetos" de conocimiento objetan y se revuelven de forma colectiva generando un conocimiento experto sobre los propios expertos (García-Dauder y Romero, 2012). *A través de la web se diversifican las voces de los diferentes actores implicados desde posiciones no homogéneas proporcionando un espacio de generación de biosocialidades y de ciudadanía digital*.

A través de la participación en una organización internacional de personas intersexuales coordinada a través de Internet, y a partir del seguimiento de un grupo de apoyo de insensibilidad a los andrógenos, y sus reuniones anuales con profesionales expertos, pudimos participar de diferentes voces y experiencias en torno a la vivencia de la intersexualidad y las relaciones con la comunidad médica y el compromiso político. También pudimos observar las posibilidades de puentes y alianzas entre las diferentes posiciones en torno a la gestión de los tratamientos, y *las dificultades de traducción entre un lenguaje médico de normalización y un lenguaje político de reconocimiento de la diversidad de género.* El reto que se nos presenta en el contexto español, es poder conjugar un discurso biomédico, en contacto directo con las personas afectadas, que pretende dar una respuesta quirúrgica y hormonal inmediata a malestares psicosociales producto de cuerpos no ajustados a normas sociales de género; con un discurso social y político más cercano a movimientos sociales globalizados relacionados con la diversidad sexual, pero lejano a la

experiencia concreta de las personas con condiciones intersexuales (en muchos casos identificadas más con el síndrome concreto que se les ha diagnosticado –SIA, Klinefelter, etc.). Nos encontramos también con *la necesidad de un espacio teórico y de intervención a caballo entre el respeto a los deseos de normalización y la crítica a la norma* (Davis, 2007). Un espacio privilegiado a su vez para analizar conflictos y negociaciones entre posiciones procedentes de diferentes disciplinas académicas (ciencias de la salud y ciencias sociales, entre otras) en articulación con otros actores implicados que también generan conocimiento experto desde su propia experiencia (personas con condiciones intersexuales, familiares, etc.).

En un plano más global, y desde un lenguaje expresamente político, diferentes colectivos intersexuales están presionando bajo el marco de los derechos humanos para que se produzcan *no solo cambios en los tratamientos médicos, sino en las propias concepciones científicas relacionadas con la identidad, el sexo y el género.* Al igual que ocurriera con el movimiento feminista o con las comunidades de enfermos de VIH, podemos estudiar las relaciones bidireccionales entre movimientos sociales y comunidad científica y hasta qué punto ciencia y política están relacionadas. Desde un plano más local, los grupos de apoyo de afectados y familiares demandan también cambios en los tratamientos. Tomando como modelo *una atención centrada en el paciente* reclaman el abandono de los silencios, las mentiras y el estigma que rodeaba al protocolo clásico propuesto por Money. Así, *se exige una comunicación abierta médico-paciente que posibilite las mejores decisiones con la mayor información*. Colectivos intersex y grupos de apoyo médicos constituyen dos actores centrales con voces diferentes en torno a la intersexualidad y con diferentes relaciones con la comunidad médica: en el primer caso, relaciones de oposición al defender la despatologización; en el segundo, relaciones de dependencia, pero también de colaboración más "paritaria" entre pacientes y especialistas (Callon y Rabeharisoa, 2003). En ambos casos, el rol de paciente-víctima de una enfermedad se transforma en un rol activo que evalúa las afirmaciones científicas, demanda cambios y genera conocimiento colectivo a partir de sus propias experiencias convirtiendo en problemática la distinción entre conocimiento experto y lego (Epstein, 1996).

Lo que queríamos estudiar en este segundo proyecto era precisamente la articulación de estos diferentes actores implicados en la gestión socio-médica de la intersexualidad (grupos de afectados y familiares, colectivos políticos, médicos, etc.) y su papel en las *controversias surgidas en torno a los tratamientos* (sus discursos, intereses, estrategias de legitimación, etc.), especialmente respecto a la eliminación de la estigmatización y las cirugías genitales tempranas "urgentes" e irreversibles. Pero en el 2006 nos encontramos con *otra controversia, esta vez en torno al cambio de nombre* (de condiciones intersexuales a DSD –*disorders of sex development*-)

constituida como "punto de paso obligado" para un cambio en los tratamientos. La particularidad es que esta controversia no surgía de forma clásica solo entre expertos científicos (como la antigua entre Money-Diamond respecto al papel determinante de la herencia/ambiente en la identidad sexual), sino entre expertos y grupos "legos" (que devienen expertos desde otros canales hasta ahora externos a la producción científica ortodoxa). De esta forma, la propuesta de cambio de nombre nos permitía analizar los efectos performativos y políticos de las categorías médicas, los movimientos y estrategias de legitimización de los diferentes actores implicados, la reconfiguración de la distinción entre saberes expertos/legos, la relación entre objetividad y democracia cognitiva, etc. (García-Dauder y Romero, 2012).

Finalmente señalar los fenómenos de "glocalización" en la recepción de estas controversias en el contexto español. El imperialismo cultural anglosajón llega a través de revistas científicas, pero también a través de foros y redes sociales creadas por grupos de apoyo médico y por colectivos políticos. Todo ello, y en diferente medida, se incorpora y muta en los contextos locales, como en el caso español.

Medios de comunicación y espacios deportivos

Otra línea de investigación que hemos considerado fundamental en la cartografía de la intersexualidad ha sido el análisis de su representación en los medios de comunicación de masas como canales de divulgación científica y generadores de opinión y percepción pública. Su estudio nos ha desplazado a otro espacio de regulación y control del dualismo sexual y, por tanto, de la intersexualidad: el deporte.

Hasta recientemente, la intersexualidad había estado prácticamente ausente de los espacios de discurso y representación de los medios de comunicación de masas. La propia palabra era o es desconocida por la opinión pública. Por ello nos propusimos investigar la representación de la intersexualidad en los MCM (fundamentalmente en la prensa nacional, pero también en otros medios). A través de este estudio, pudimos observar cómo en la prensa escrita sigue presente el imaginario del *hermafrodita como monstruo moral* del siglo XIX. Junto a ello, la otra principal representación es la de la *anomalía clínica*, el referente médico que reduce la persona al cuerpo fragmentado "sexualmente ambiguo, indefinido" que debe ser corregido hormonal y quirúrgicamente para ser "normal"; por otro lado, como señalábamos, objeto de interés científico y recurso para dirimir controversias herencia-ambiente, naturaleza-cultura, etc. Por otro lado, la intersexualidad como "anomalía o patología física" se representa en oposición a la transexualidad como "alteración mental". Con ello el mensaje enviado a la opinión pública es la justificación de las cirugías genitales como una "necesidad médica" en el primer caso, y la no cobertura pública

sanitaria en el segundo consideradas como cirugías estéticas. Por último, la estigmatización y la cosificación vienen acompañadas en los medios de la *victimización*, cuando se narran experiencias vividas, las personas intersexuales aparecen como víctimas pasivas bajo el sufrimiento inescapable de ser diferente. En definitiva, en la mayoría de los casos, las representaciones sobre la intersexualidad en estos medios interpelan a la persona intersexual por medio de palabras e imágenes que estigmatizan, cosifican o victimizan, con los efectos sociales y subjetivos que ello conlleva (García Dauder, en prensa).

En los últimos años la intersexualidad ha saltado a la opinión pública a partir de noticias donde *se ponía el sexo de deportistas mujeres bajo sospecha*. A partir de este material derivamos al análisis de la construcción del sexo mujer, la regulación del dualismo sexual y el control de la intersexualidad en el deporte a partir de los "tests de verificación de género". Si bien estos tests se crearon en un primer momento para detectar posibles fraudes de varones que se hicieran pasar por mujeres; posteriormente, se han convertido en reguladores de "casos de intersexualidad" en el deporte, con el efecto de no permitir competir a deportistas cuya biología o naturaleza no respondía a los estándares considerados por el ámbito deportivo "aceptables" para una mujer. De nuevo, nuestro objetivo ha sido analizar bajo qué criterios se define a una mujer en el deporte y cuáles son los presupuestos socio-culturales y valores de género implícitos en ello. Si en los años 60 el análisis se determinaba a partir de la inspección anatómica externa, y en los 80-90 a partir del análisis cromosómico y genético, hemos podido observar cómo se han desplazado los criterios determinantes sobre la base de qué componentes confieren una "ventaja competitiva" a las deportistas. Y así hoy en día es el criterio hormonal, "los niveles aceptables de testosterona en una mujer", el que determina si una deportista podrá competir o no *como mujer.* Del control a que participen "mujeres verdaderas" (mediante pruebas anatómicas y cromosómicas) se ha pasado a la regulación de la "ventaja competitiva" (fuerza muscular y niveles aceptables de "hormonas masculinas" generados por el propio cuerpo). De esta forma se asegura la homogeneidad de los cuerpos sexuados de las mujeres a la baja: el no tener una ventaja competitiva te convierte en "mujer" y a la inversa. Al igual que en el espacio médico, o en los medios de comunicación, nuestro objetivo ha sido investigar las versiones de la intersexualidad, y con ello del dualismo sexual, en el ámbito deportivo (García Dauder, 2011).

De ciudadanía sexual, bioética y gobernanza: líneas futuras

Finalmente señalar algunas líneas a las que apuntaríamos en un posible proyecto de investigación futuro. En la investigación biomédica de la intersexualidad/DSD, nos encontramos ante un nuevo escenario de "avances científicos" que ya no se orienta tanto a las tecnologías

quirúrgicas u hormonales sobre cuerpos ya nacidos, sino a la investigación en biología molecular de la determinación sexual y sus "alteraciones". Como ha señalado Paula Machado (2009) se incorpora el "sexo-código", un sexo que se revela a nivel molecular del cuerpo, en su dimensión más invisible y menos controlable por la propia persona. Siguiendo a Rose (2007), la gestión de los "cuerpos intersexuales" ya no sólo se dirime en un nivel de análisis anatómico-molar pos-natal (quirúrgico, endocrino o psicológico), sino que se pretende situar en un nuevo nivel genético-molecular, que a su vez viene avalado por las nuevas "tecnologías de optimización" de diagnóstico e intervención pre-natal. Ello nos coloca en un nuevo escenario bioético ante las posibilidades de que esta biopolítica molecular con sus tecnologías delinee una ciudadanía sexual deseable y vivible.

Otras líneas futuras que entroncan con consideraciones bioéticas tienen que ver con el consentimiento informado, la autonomía del paciente, las decisiones de los padres o familiares y el mejor interés para los niños. Todo ello se pone en juego (y en muchos casos de forma incompatible) en las intervenciones médicas a bebés intersexuales (fundamentalmente cuando hablamos de cirugías genitales tempranas e irreversibles).

Para terminar, dentro del proyecto más amplio "Cartografías del cuerpo", continuaremos investigando la confluencia de análisis procedentes de diferentes tecnologías biomédicas y su comparación. Nos interesa identificar con ello diferentes modelos de gobernanza y patrones de intercambio de saberes y relaciones expertos/legos; diferentes biosocialidades y su participación pública; y el papel de Internet en el marco de una nueva ciudadanía digital y de la salud. A su vez, continuaremos analizando el "sexo múltiple" en sus diferentes versiones, y cómo se conforma médicamente la normalidad de género y de los cuerpos sexuados. Todo ello, por último, desde la reflexión epistemológica sobre cómo nos movemos respecto a un "objeto de estudio"; las relaciones entre ciencia y política, investigación y activismo, especialmente cuando se trata de temas de sexo/género; o los problemas de las portavoces, del "hablar por", y la necesidad de políticas de articulación no estáticas ni rígidas.

REFERENCIAS

Cabral, Mauro (2009). No saber –acerca de XXY. En Mauro Cabral (ed.), *Interdicciones. Escrituras de la intersexualidad en castellano* (105-09). Córdoba: Anarrés ed.-Mulabi.

Callon, Michel y Rabeharisoa, V. (2003). "Research 'in the wild' and the shaping of new social identities". *Technology in Society,* 25, 193-204.

Chase, Cheryl (1998). Hermaphrodites with attitude. *Journal of lesbian and gay studies, 4* (2): 189-211.

Davis, Kathy (2007). *El cuerpo a la carta. Estudios culturales sobre cirugía cosmética.* México: La cifra.

Epstein, Steven (1996). *Impure Science*: *AIDS, Activism, and the Politics of Knowledge.* Berkeley: University of California Press.

García Dauder, Silvia (2011), Las fronteras del sexo en el deporte: tecnologías, cuerpos sexuados y diferencias, *INTERthesis. Revista Internacional Interdisciplinar, 8(2),1-19.*

García Dauder, Silvia (en prensa). La regulación tecnológica del dualismo sexual y el diseño de cuerpos normativos. En E. Pérez Sedeño y E. Ortega (eds.), *Los cuerpos de la ciencia.* Madrid, Cátedra.

García Dauder, Silvia y Romero Bachiller, Carmen (2012). Los desplazamientos políticos de las categorías médicas. En E. Pérez Sedeño e Ibáñez Martín, R. (ed.), Cuerpos y diferencias (213-240). Madrid: Plaza y Valdés.

García Dauder, Silvia; Romero Bachiller, Carmen y Ortega, Esther (2007). Cossos insubmissos i mutilació genital: La regulació tecno-mèdica dels cossos sexuats. En J.M. Armengol (coord.), *Masculinitats per al segle XXI* (35-45). Barcelona: CEDIC.

Machado, Paula Sandrine (2009): "Confesiones corporales: Algunas narrativas socio médicas sobre los cuerpos intersex", en Mauro Cabral (ed.), *Interdicciones. Escrituras de la intersexualidad en castellano*, Córdoba: Anarrés ed.-Mulabi, pp.83-100.

Rose, Nikolas (2007): *The politics of life itself. Biomedicine, power and subjectivity in the twenty-first century,* Princeton and Oxford: Princeton U.P.

SOBRE LA MEDICINA ESTÉTICA Y LA INEVITABILIDAD DE LA TECNOLOGÍA

EULALIA PÉREZ SEDEÑO
Dpto. Ciencia, Tecnología y Sociedad
IFS-CSIC

ABSTRACT: Cosmetic Medicine and Inevitability of Technology
Cosmetic medicine practice is a medical intervention aimed at enhancing or improving healthy people's bodies. Nowadays, its practice is so widespread that it is one of the items calculated in the Spanish consumer price index. The rise of cosmetic surgery is due to diverse factors such as the introduction of anesthesia and antibiotics, and to social, cultural and economic factors. In recent years, it has experienced a far-reaching diffusion greater than any other biomedical technology. Despite strong initial resistance especially from some sectors, cosmetic medicine, especially surgery, has introduced an unprecedented medical practice which seems to follow the technological imperative: since it can be done and is feasible, it *should* be done. This expansion of cosmetic surgery and medicine might teach us a lesson applicable to other scientific-technological practices, such as cosmetic neurology or genetic engineering for aesthetical purposes.
KEY WORDS: cosmetic medicine, cosmetics genetics, ethics and technology, gender.

Introducción

La medicina estética es una práctica médica que mejora a la gente sana y está tan extendida que desde enero de 2007, es uno de los ítems que se computan en el índice de precios al consumo (IPC) español. Su práctica es sumamente antigua, pues disponemos de descripciones de correcciones nasales y tratamientos de cicatrices ya en el antiguo Egipto. Con la aparición de la sífilis en Europa en el siglo XV, que desfiguraba y estigmatizaba a los enfermos, se empezaron a desarrollar técnicas de restauración, demandadas en este caso por los hombres.

La introducción de la anestesia (1846) y los antibióticos (1897 por Ernest Duchesnes, aunque la terapéutica antibiótica moderna comenzó con Paul Ehrlich en 1909) desempeñaron un importante papel a la hora de que estos procedimientos quirúrgicos fueran más viables. La cirugía estética, tal y como la conocemos ahora, se origina, no obstante, en la Primera Guerra Mundial, en concreto en Alemania, debido a la gran cantidad de víctimas y heridos. Luego, durante años, la cirugía estética estuvo reservada a las clases altas y a las personas que se dedicaban a la industria cinematográfica, sobre todo desde que en los años 20 del pasado siglo se popularizara el "primer plano", muy aplaudido por el público, pero que revelaba el más mínimo defecto o imperfección. Y en los últimos años, ha experimentado una difusión mayor que ninguna otra tecnología biomédica. ¿Cómo ha llegado a ser tan común?

'Norma' y tecnología

Para el desarrollo de la cirugía estética era necesaria una serie de prerrequisitos técnicos y científicos. Pero en su auge intervienen otros factores culturales, sociales y económicos (Pérez Sedeño, 2012). Por ejemplo, en el siglo XX se pasó de hacer hincapié en el carácter del individuo, al culto a la personalidad, enraizado en el juicio social más que en el individual: vestirse y hablar de manera adecuada y tener buenos modales eran considerados indicativos de una buena personalidad. También los cambios en el lugar de trabajo (la población pasó de entornos rurales aislados a entornos urbanos más competitivos y transitorios) tuvieron su importancia, pues la primera impresión que una persona daba se convirtió en algo importante. Así pasó a considerarse que en lugares de trabajo sumamente competitivos, la decisión de mejorar la propia apariencia era una decisión racional y económica. Este argumento se utilizó para justificar el uso de la cirugía reconstructiva con fines puramente estéticos. Por otro lado, la psicología proporcionó un marco conceptual a la práctica de la medicina estética: la idea de que una persona pudiera desarrollar un "complejo de inferioridad" debido a imperfecciones físicas, que disminuyera su autoestima y afectara su trabajo, proporcionó una justificación más. De este modo, los cirujanos estéticos dejaron a un lado el temor de que se les acusara de fomentar la frivolidad (Pérez Sedeño, 2012).

La práctica de las tecnologías estéticas planteó algunos problemas desde el inicio de su expansión, a partir de los años ochenta. Por un lado, han surgido dudas acerca de la seguridad de sus procedimientos quirúrgicos. En situaciones de enfermedad, sopesamos los riesgos potenciales con respecto a los beneficios potenciales, pero cuando no hay enfermedad debemos preguntarnos qué riesgos merecen la pena ser tomados. También se han planteado cuestiones sobre la calidad del consentimiento informado o sobre la certificación y cualificación de los cirujanos. En tercer lugar, aunque desde la ética apenas se le ha prestado atención, se plantea una cuestión sobre justicia distributiva: si la medicina estética hace "mejores" o más bellas a las personas, ¿estarán disponibles de forma justa estas mejoras para toda la gente? Por lo general, los sistemas nacionales de salud cubren la cirugía reparadora[4] y algunas cirugías que corrigen perturbaciones funcionales, como la caída del párpado que dificulta la visión. Sin embargo la cirugía estética no está incluida entre las prestaciones sanitarias y es de pago[5]. Finalmente, está también la cuestión de la coerción: ¿en qué medida las personas que se someten a estas prácticas científico-tecnológicas las eligen libremente?

[4] La cirugía plástica se divide en reparadora y estética. El objetivo de la primera es restaurar o mejorar la función y el aspecto físico en piel o partes del cuerpo, que han sufrido lesiones causadas por accidentes, quemaduras o enfermedades. La segunda, en cambio, se ocupa de pacientes sanos para alterar su fisionomía en diferentes aspectos y adecuarlas a ciertas normas estéticas (Sociedad Española de Cirugía Plástica, Reparadora y Estética (SECPRE): http://www.secpre.org/index.html. consultada el 1 de mayo de 2012).

[5] Algunos sistemas de salud, como el holandés, la incluyen cuando la condición encaja en uno de los siguientes tres criterios: crea una perturbación funcional (por ejemplo la caída del párpado); provoca un grave sufrimiento psicológico; o es una imperfección física que queda fuera del grado normal de variación de la apariencia (Davies, 1995, 2003). Como vemos, conceptos como 'sufrimiento psicológico' o 'normal' son, cuando menos, vagos.

No cabe duda del papel desempeñado por los medios de comunicación en la aceptación de unas prácticas que pretenden adaptar los cuerpos a las normas sociales y culturales de lo que *debe ser* un cuerpo 'femenino' y tienden a tratarlas de forma positiva sin apenas mencionar los riesgos, si es que lo hacen.

Algunas autoras argumentan que la cirugía estética exacerba "concepciones dañinas de normalidad" (Little, 1998), mediante unas normas de apariencia que se dirigen principalmente a las mujeres y que especifican cómo debe ser su apariencia física y para lograrla se les exige una importante inversión de tiempo, energía y dinero (Pérez Sedeño, 2012a). Como la mayoría de las mujeres normales no encajan con el ideal, incluso las que tienen cuerpos sanos creen que tienen "deficiencias" estéticas y se sienten insatisfechas con sus cuerpos sin cuestionarse la presión social que sobre ellas se ejerce para ajustarse a estas "normas".

Desde el feminismo se considera, por lo general, que las presiones para tener una apariencia joven, delgada, de piel suave, ojos grandes, etc. están enraizadas en injusticias históricas, pero no hay opinión unánime acerca de cómo entender el papel de los individuos en la popularidad de la cirugía estética. Cómo se entiende la relación entre los deseos y motivaciones del individuo y los dictados de la sociedad lleva a diferentes estrategias para abordar el problema de la presión para ajustarse a las normas de belleza.

Según una de ellas, las mujeres que se someten a cirugía estética siempre lo harían *debido a* normas dañinas, a pesar de que afirmen lo contrario. Esta idea presenta a las mujeres como "idiotas culturales" controladas por su entorno e inconscientes de ese control. "Las mujeres que se han sometido a cirugía estética o que están pensando someterse a ella niegan sistemáticamente la influencia de las imágenes de los medios. 'Lo hago por mi' insisten. Pero es difícil explicar la mayoría de sus acciones (las operaciones que se efectúan más frecuentemente son el aumento de pecho y la liposucción) fuera del contexto de las normas culturales actuales" (Bordo, 1998). No obstante, aunque reconoce el papel de las influencias culturales, Bordo no considera que las mujeres sean idiotas culturales, que no tengan ningún sentido de agencia. Piensa que, aunque a las mujeres les influyen las normas culturales también deben reconocer su propia complicidad "en la perpetuación de normas racializadas y en el sufrimiento de otras personas"(*ibídem*) y en cómo sus acciones aumentan la presión para encajar en la norma.

Al participar en la cirugía estética, estas mujeres se alejan de las realidades del envejecimiento y el cambio porque la sociedad no considera atractivos los rasgos asociados con la edad. Quieren evitar *ser ellas mismas*, pero afirman que lo hacen *por ellas mismas*. En respuesta a esas mujeres que afirman haber descubierto finalmente sus yos reales a través de la cirugía estética (una afirmación que plantea cuestiones interesantes sobre la autenticidad y la identidad), Bordo

insiste en que *se niegan* a sí mismas la oportunidad de entender nuestra condición humana compartida de vulnerabilidad física, mortalidad y permanencia y también *refuerzan* concepciones dañinas de normalidad a través de sus acciones. En efecto sus acciones aumentan la presión para adecuarse a la norma.

Pero si las mujeres que eligen la cirugía estética son simples idiotas culturales, parece que hay que absolverlas de responsabilidad por sus acciones. Simplemente siguen la dirección de fuerzas externas que *conforman sus* deseos. La mejor solución para las concepciones dañinas de normalidad que acepta la idea de la "idiota cultural" es cambiar las presiones sociales. Esto podría lograrse, por ejemplo, exigiendo que la industria publicitaria presente más diversidad en las formas corporales de las modelos. En este sentido ha habido varios intentos como el *Body Image Summit* (June 2000) que reunió a representantes del gobierno británico con editores de moda para ver cómo manejar las imágenes de la publicidad y asegurar que habría diversidad en la forma y el tamaño de las modelos; o el *Real Women Project* un esfuerzo del sur de California que emplea el instrumental de los anuncios (imágenes, música, literatura) para promocionar la auto-aceptación y apelar al sentido común de las mujeres[6]. En España, ha habido una iniciativa con las pasarelas de moda, que cristalizó en la prohibición de desfilar a modelos con un índice de masa corporal[7] inferior al considerado saludable, y que condujo, en 2006, a rechazar al 30% de las modelos que se presentaron para desfilar en la pasarela Cibeles por no alcanzar un peso saludable[8] . También hubo un estudio antropométrico de la población femenina realizado por el Ministerio de Sanidad, presentado en 2008, para determinar las formas típicas de los cuerpos de las españolas, que pretendía servir como punto de partida para elaborar un nuevo patronaje para el sector textil y que las mujeres no se sintieran 'fuera de la norma'.

El discurso que usan los medios (incluidas las páginas webs informativas de las clínicas de estética), así como los propios médicos, convierten estas prácticas en algo familiar y ocultan lo que en realidad son: la medicalización absoluta de los cuerpos. Las modificaciones estéticas de nariz, mentón u orejas se convierten en tratamientos de *desgracias genéticas*; la eliminación de *michelines*, realzado o rellenado de senos o torneado de muslos y nalgas se convierten en tratamientos de los *daños* ocasionados por el embarazo o la edad; los peelings químicos o mediante laser pasan a ser tratamientos de los *daños* causados por el sol, el tabaco y el acné.

En realidad, la cirugía estética no sólo medicaliza los cuerpos, sino que los redefine como objetos de reconstrucción tecnológica. Produce un significado cultural en el que podemos examinar

[6] http://www.realwomenproject.com/. Último acceso, 25 de mayo de 2012.

[7] El índice de masa corporal relaciona el peso de cada persona con su estatura y permite conocer con carácter general su estado nutricional. Su cálculo se realiza al dividir el peso en kilos por el cuadrado de la estatura, y el resultado debe oscilar entre 18 y 25 para considerarse saludable.

[8] *El País,* 7 de Septiembre de 2006.

la reproducción literal y material de los ideales de belleza. Los procedimientos quirúrgicos cortan la carne y aíslan las partes para manipularlas y re-esculpirlas. De ese modo, la cirugía estética literalmente transforma el cuerpo material en un signo cultural. Y no es simplemente un lugar discursivo donde "se construyen las imágenes de las mujeres, sino un lugar material en el que el cuerpo femenino físico se disecciona, se estira, se corta y se reconstruye según normas sumamente ideológicas de apariencia física" (Balsamo, 1999).

A pesar de la enorme resistencia inicial, sobre todo desde algunos sectores, la medicina estética, en especial la cirugía, ha introducido una práctica médica sin precedentes que parece seguir el imperativo tecnológico: como puede hacerse y es factible *debe* hacerse. ¿Puede esta expansión de la cirugía y medicina estética enseñarnos alguna lección para otras prácticas científico-tecnológicas como la neurología estética o la ingeniería genética con fines estéticos?

Consideraciones finales

Algunas personas anticipan un gran mercado para las técnicas genéticas que se apliquen para la mejora estética. Si podemos someternos a cirugía para cambiar nuestros cuerpos según nuestros deseos, ¿por qué no elegir terapias genéticas que produzcan cambios corporales para uno mismo o para los futuros hijos (por ejemplo seleccionando embriones ya tratados con fines estéticos e implantarlos despues)? En la genética estética el propio cuerpo produce los rasgos deseados como por ejemplo tener ojos azules, ser alto, mantener una proporción baja de grasa corporal, desarrollar pechos grandes, o parecer menos "étnico" diseñando la forma de la nariz o del ojo, la textura del pelo, el color de la piel, etc.

Los debates sobre la ética de la cirugía estética pueden informar el debate futuro sobre si es apropiada la genética estética. Pero incluso antes de que ésta se convierta en una realidad, los ciudadanos pueden reconocer los peligros de la genética estética y actuar para promover legislaciones, distribuir los fondos para la investigación de determinado modo e iniciar programas de educación para todo el público. La prudencia sugiere situar una moratoria temporal sobre la financiación pública de la investigación genética diseñada para identificar u ofrecer terapia para rasgos primariamente estéticos, pero no así para los desórdenes genéticos devastadores o severos. Sin embargo, determinar qué desórdenes son 'severos' no es fácil: el movimiento por los derechos de los discapacitados o diversos funcionales observa que muchas condiciones que el público general estaría tentado de "curar" mediante la ingeniería genética, pueden no ser considerados fisiológicamente indeseables. Si finalmente se materializan esas pruebas de diagnóstico específico o las terapias genéticas, los hospitales públicos podrían imponer regulaciones sobre el uso de esas

pruebas. ¿Es eso aceptable? ¿Quién o quienes deberían regular y cuáles son los mecanismos adecuados para ello?

Aunque los hospitales impongan sus propias restricciones, parece probable que los empresarios den un paso adelante y ofrezcan esos servicios fuera de los establecimientos médicos regulares y de los sistemas de salud, siguiendo las 'leyes del mercado' (como en el caso de la cirugía estética). En ese caso podríamos abogar por un programa amplio de educación pública, aduciendo que es el mejor modo de combatir esa demanda. Los programas de educación pública que subrayan la salud y la belleza y singularidad de diversas formas corporales, nos ayudarían a estar más satisfechos con nuestros cuerpos (y aceptar un futuro hijo que no encaje con el ideal). Con sincero esfuerzo, podríamos ser capaces de abandonar un ideal basado en un tipo corporal físico específico y abrazar un ideal que haga hincapié en compromisos tan profundos como la participación en la sociedad, el valor intelectual y el dar cuidado emocional. Una mejor financiación de programas que se centren en estos compromisos más profundos podría acelerar el cambio. Por ejemplo, *Girls Incorporated* es un programa que pretende ayudar a las jóvenes "a enfrentar mensajes sociales sutiles sobre su valor y potencial". Dentro del programa hay un acta de derechos que subraya el "derecho aceptar y disfrutar los cuerpos con los que las chicas nacieron y no sentirse presionadas a comprometer su salud para satisfacer los dictados de una imagen física "ideal"[9].

La genética estética puede aprender varias lecciones de la cirugía estética: si una práctica contribuye a, o refuerza, concepciones dañinas de normalidad que deben cambiarse, hay que buscar otros medios de lograr los intereses individuales, pues a veces desechamos alternativas demasiado rápidamente (si otra persona lo hace, podríamos quedar en desventaja). Además, si nos fijamos en la historia de la cirugía estética, el nacimiento y desarrollo de la genética estética parece inevitable: los factores que impedían el desarrollo de la cirugía estética han desaparecido, mientras que los factores que la promovían siguen operando. La genética estética plantea serios problemas y retos a la medicina y a la bioética. La bioética surgió de forma institucionalizada a finales de los años 80 y principios de los 70 del pasado siglo y se centró en los avances médicos y tecnológicos. La cirugía estética surgió y se desarrolló bastante antes, sin que los bioéticos profesionales le prestaran casi atención y poco análisis crítico. Aunque no ha sucedido así con la genética, ¿tendrá la bioética algún influjo sobre la genética estética?

REFERENCIAS

[9] http://www.girlsinc.org/. Último acceso, 19 de mayo de 2012.

Balsamo, Anne (1999): *Technologies of the gendered body*, Durham-Londres, Duke University Press.

Bordo, Susan (1998): "Braveheart, Babe, and the Contemporary Body" en Parens ed.

Davis, Kathy, (2003): *Dubious Equalities and Embodied Differences. Cultural Studies on Cosmetic Surgery*, Rowman & Littlefield Publishers.

Little, Margaret (1998): "Cosmetic Surgery, Suspect Norms, and the Ethics of Complicity" en Parens ed.

Parens, Erik (ed.) (1998): *Enhancing Human Traits Ethical and Social Implications,* Washington DC: Georgetown University Press,

Pérez Sedeño, Eulalia (2012): "Ciencias y tecnologías del cuerpo: la práctica de la tecnología de los implantes mamarios", en Pérez Sedeño, E. e Ibáñez Martín, R. *Cuerpos y diferencias*, Madrid-México, Plaza y Valdés

(2012a): "¡Una talla 100, por favor!" en Pérez Sedeño, E., Juan Carlos Toscano Grimaldi, Consuelo Flecha García, Ana Guil Bozal, Óscar Macías Álvarez y Mª Encarnación Pinero Fernández (eds.): *Memorias del IX Congreso Iberoamericano de Ciencia, Tecnología y Género*. Disponible en http://www.oei.es/congresoctg/memoria/

C U R S O S E C O N G R E S O S
Nº 2 1 7

SCIENTIFIC COMMITEE/ COMITÉ CIENTÍFICO

Juan José Acero (Universidad de Granada) • John Corcoran (Buffalo University, USA) • Javier Echeverría (Ikerbasque/ Universidad del País Vasco) • Manuel García-Carpintero (Universidad de Barcelona) • María Manzano, Grzegorz Malinowski (Universidad de Lódz, Polonia) • C. Ulises Moulines (Ludwig-Maximilians-Universität München, Alemania) • León Olivé Ordoñez (Universidad Autónoma de Madrid) • Stewart Shapiro (Ohio University, USA -St Andrews University, UK)

ORGANISING COMMITTEE/ COMITÉ ORGANIZADOR

Concha Martínez Vidal (Presidenta del Congreso, USC) • José Luis Falguera López (Secretario, USC) • Xavier de Donato (Vocal, USC) • José Miguel Sagüillo Fernández-Vega (Vocal, USC) • María José Frápolli Sanz (Vocal, Universidad de Granada, Presidenta de la SLMFCE) • Pierdaniele Giaretta (Vocal, Universidad degli Studi di Padova, Italia) • Stephen McLeod (Vocal, University of Liverpool, UK) • Sofía Miguens (Vocal, Universidade de Porto, Portugal) • Martín Pérez Pereira (Vocal, USC) • Uxía Rivas Monroy (Vocal, USC) • Alejandro Sobrino Cerdeiriña (Vocal, USC) • Charles Travis (Vocal, King's College, UK/ Univerdidad de Porto) Juan Vázquez Sánchez (Vocal, USC) • Jesús Vega Encabo (Vocal, Universidad Autónoma de Madrid) • Víctor Verdejo Aparicio (Vocal, USC) • Luis Villegas (Vocal, USC) • Inés Pan Espiñeira (Vocal, USC)

First edition Lullius lectures / Primera edición conferencias Lullius
PHILIP KITCHER

Proyectos colaboradores
FFI2009-08828/FISO, FFI2011-13481-E

Sociedad de Lógica, Metodología
y Filosofía de la Ciencia

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA