VI CONGRESO DE LA SOCIEDAD DE LÓGICA, METODOLOGÍA Y FILOSOFÍA DE LA CIENCIA EN ESPAÑA (SLMFCE)

JESÚS ALCOLEA
VALERIANO IRANZO
ANA SÁNCHEZ
JORDI VALOR
(**Editores**)

ACTAS DEL VI CONGRESO DE LA SOCIEDAD DE LÓGICA, METODOLOGÍA Y FILOSOFÍA DE LA CIENCIA EN ESPAÑA (SLMFCE)

18 - 21 de noviembre

VALENCIA
2009

Diseño de cubierta: Celso Hernández de la Figuera

ISBN: 978-84-370-7655-3
Depósito legal: V-XXXX-2009

Impresión: Diazotec, S. A.

COMITÉ CIENTÍFICO

Atocha Aliseda Llera (Universidad Nacional Autónoma de México)
Theo A. F. Kuipers (University of Groningen, The Netherlands)
Alberto Moretti (Universidad de Buenos Aires, CONICET)
Carlos U. Moulines (Ludwig-Maximilians-Universität München)
François Recanati (Institute Jean Nicod, París)
Gabriel Sandu (CNRS, París)

COMITÉ ORGANIZADOR

Jesús Alcolea Banegas (UVEG)
Valeriano Iranzo García (UVEG)
Ana Sánchez Torres (UVEG)
Jordi Valor Abad (UVEG)
María José Frápolli Sanz (UGR, Presidenta de la SLMFCE)
Vicente Claramonte Sanz (UVEG)
Jesús Vega Encabo (UAM, Secretario de la SLMFCE)
David Porcel Muñoz (UVEG)
Rafael Beneyto Torres (UVEG)
Enric Casaban Moya (UVEG)
José Pedro Úbeda Rives (UVEG)
Eulalia Pérez Sedeño (IFS, CCHS, CSIC)

PONENTES INVITADOS

Helen Longino (Stanford University, Estados Unidos)
John Worrall (London School of Economics, Reino Unido)
Johan van Benthem (Universidad de Amsterdam, Países Bajos)

INSTITUCIONES PATROCINADORAS

Ministerio de Ciencia e Innovación
Vicerectorat d'Investigació i Política Científica (Universitat de València)
Càtedra de Divulgació de la Ciencia (Universitat de València
Facultat de Filosofia i Ciències de l'Educació (Universitat de València)
CAM (Caja de Ahorros del Mediterráneo)
Departament de Lògica i Filosofia de la Ciència (Universitat de València)

INSTITUCIONES COLABORADORAS

Bancaixa (Caja de Ahorros de Valencia, Alicante y Castellón)
Agència Valenciana del Turisme. Generalitat Valenciana

Prólogo

Este libro contiene las comunicaciones aceptadas en el VI Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España celebrado en Valencia del 18 al 21 de noviembre de 2009. El volumen contiene 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 los estudios sobre ciencia, tecnología y sociedad. Hemos de señalar que todas las propuestas recibidas fueron sometidas a un proceso de "recensión ciega" que involucró un buen número de académicos e investigadores. El porcentaje de aceptación global ha sido ligeramente superior al 80 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.

En la línea de lo acontecido en las últimas ediciones de este congreso, nos congratulamos por la presencia significativa de comunicaciones procedentes de otros países europeos y latinoamericanos, tanto entre los participantes a título individual como en las mesas y simposios celebrados en el seno del congreso. Ello es un buen indicador del potencial de estos eventos para abrir y preservar cauces de comunicación entre investigadores de diferentes países.

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. También queremos mencionar al resto de miembros del comité organizador, y a quienes han colaborado en el proceso de recensión 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 y mesas.

Valencia, octubre de 2009

Premio para Jóvenes Investigadores
VI congreso de la SLMFCE

Con motivo de la VI 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 tres premios, consistentes en bolsas de viaje, para las mejores contribuciones al congreso a cargo de investigadores jóvenes.

Las ponencias ganadoras fueron:

'What Does Embodiment Mean? Questioning the Autonomy of Psychology'
Saray Ayala López

'Demonstrating Fictional Names'
Gemma Celestino Fernández

'Paraconsistent Vagueness: A Positive Argument'
Pablo Cobreros

Índice

COMUNICACIONES

Sección A. Lógica, historia y filosofía de la lógica

Sección B. Filosofía del lenguaje, filosofía de la mente, epistemología

Sección C. Filosofía y metodología de la ciencia

Sección E. Ciencia, tecnología y sociedad

SIMPOSIA

Simposio «Argumentación»

Mesa redonda «Importancia de los conceptos»

Sección A

Lógica, historia y filosofía de la lógica

Un *input* para la máquina universal de Turing

Rafael Beneyto
Universitat de València
rafael.beneyto@uv.es

Introducción

Lo que identifica a una máquina de Turing y la diferencia de todas las demás está constituido por el conjunto de sus estados, el conjunto de sus símbolos y las funciones que determinan sus operaciones a partir del estado en que se encuentra y el símbolo que escudriña. En este trabajo procuraremos hacer irrelevantes la naturaleza de sus diversos estados y el número y características de sus símbolos en el siguiente sentido: sin importar de qué máquina de Turing se trate podemos definir otra equivalente cuyas computaciones están determinadas por

a) el número de sus diferentes estados y la identificación de los estados finales; y

b) las funciones que especifican sus operaciones.

Sabemos que para que una máquina de Turing universal pueda realizar el cómputo que corresponde a otra máquina de Turing (ella misma incluida) esta máquina ha de proporcionársele como input debidamente codificada en términos de la simbología que baraja la primera máquina. Presentaremos una especie de máquinas que sólo operan con los símbolos "1" y "0", además del (pseudo)-símbolo "B" para el blanco. Los *n* estados propios de cada una de ellas son los *n* primeros números naturales {1, 2,..., *i*-1, *i*, *i*+1,..., *n-1, n*}. El estado inicial es siempre el "1". Las máquinas de esta especie estarán en condiciones de que su conducta sea simulada por la máquina universal de Turing presentada por Hoptcroft-Ullman. Y como ésta sólo admite "1", "0" y "B", necesitamos que también sus diversos símbolos sean de alguna manera llevados a éstos últimos.

Ofrecemos un procedimiento que, para cualquier máquina, arrojará como resultado una máquina idónea para ser input de la máquina de Hoptcroft-Ullman. Es decir:

Dada una máquina de Turing M existe una máquina de Turing M' equivalente tal que

1) su estado inicial es "1"; y

2) los símbolos que utiliza son "1", "0" y el pseudosímbolo "B".

Codificación de símbolos y cadenas de símbolos

Como los distintos símbolos de la máquina dada *M* serán distintos de "1", "0" y "B" (e incluso los superen en número), necesitamos una función de traducción *T* tal que

a) si σ es un símbolo de *M* entonces σ_T es su traducción en *M'* (donde σ_T es una cadena se símbolos "1" o "0");

b) si θ es una cadena input de *M* y θ' es la correspondiente cadena output de *M* y si θ_T es la traducción de θ y Γ es la correspondiente cadena output de *M'* entonces Γ es θ'_T.

Esto es,

$\theta \vdash_M \theta'$ si $\theta_T \vdash_{M'} \theta'_T$ (donde "$\vdash_M$" significa "determina en *M* un output")

(Dicho informalmente, si el input θ determina en *M* un output θ' entonces la traducción θ_T de aquel input determina en *M'* un output θ'_T que es la traducción de aquel primer output; y viceversa.)

Si Σ es el conjunto de símbolos de *M* la función de traducción *T* responde a tres casos: Si Σ contiene un solo símbolo, se hace corresponder "0" (alternativamente, "1") con el único símbolo; si contiene dos símbolos se hace corresponder "0" y "1" (alternativamente "1" y "0") con el primer y el segundo símbolo respectivamente.

Pero si $\Sigma = \{\sigma_0, \sigma_1,...,\sigma_n, B\}$ (para $n > 1$) entonces habremos de recurrir a cadenas de símbolos de *M'* de unos y ceros que, por conveniencia operativa, tendrán todas una misma longitud *z,* donde en concreto

$$z = \begin{cases} \log_2 (n+1), \text{ si } \mathrm{Int}(\log_2 (n+1)) = \log_2 (n+1) \\ \mathrm{Int} (\log_2 (n+1)) + 1, \text{ en caso contrario,} \end{cases}$$

Cada una de estas cadenas será la notación binaria de uno de los 2^z primeros números naturales.

Sea $[i]^z$ la cadena de *z* elementos en notación binaria del número *i*. Nuestra función de traducción *T* es tal que

$(\sigma_0)_T = [0]^z$
$(\sigma_1)_T = [1]^z$
$(\sigma_2)_T = [2]^z$
.
. y, en general,
$(\sigma_i)_T = [i]^z$
(mientras que trivialmente $(B)_T$ = BBB... *z* veces)

Dicho vulgarmente, es suficiente asignar a "B" una cadena de *z símbolos "*B*"*, ordenar el resto de símbolos y asignar a cada uno de ellos su número de orden menos 1 en binario con *z símbolos.*

Diseño de la estrategia a desarrollar

La máquina *M'* a definir opera en cada movimiento teniendo en consideración no una cadena de *z* símbolos, sino uno solo de ellos: un "1", o un "0" (o un "B"). El propósito que nos debemos plantear es que *M'* actúe de tal modo que

1º) si se le ofrece un input que no es traducción de un input de *M* entonces *M'* se detiene sin alcanzar un estado final;

2º) en el caso de que lo sea, la computación que *M'* realice en cada bloque de *z* símbolos dé como resultado la traducción del cómputo operado por *M* sobre el símbolo cuya codificación es dicho bloque;

3º) que el cómputo final de *M'* para dicho input sea la traducción del output de *M* para el input de que es codificación; y

4º) que *M'* se detenga en un estado final sólo si *M* también lo hace para el input de que se parte.

Los objetivos a alcanzar con *M'* son: lectura e identificación del bloque de *z* símbolos traducción del símbolo escudriñado por *M*, escritura del bloque de z símbolos traducción del símbolo que imprimiría *M*, desplazamiento a izquierda o derecha de *z* casillas y alcanzar el estado que se corresponde con el estado a que mutaría *M.*

Estos objetivos determinarán las funciones para *M'* a partir de las funciones de *M*. Utilizaremos los estados como almacén de datos.

Estados de M'

Además de todos los estados de *M* necesitaremos para *M'* un número considerable de nuevos estados de lectura y de escritura, que habremos de definir cuidadosamente.

Si, por ejemplo, "101" es la traducción de un símbolo de *M* y "η" un estado de ésta, podemos contar en *M'* con este estado y con los estados η_1, η_{10},... con el fin de controlar cuál es el fragmento de dicha traducción que ha sido leído a partir del estado η. Con el estado η_{10} se controla la situación de que, partiendo del estado η, la cabeza lectora ha detectado primero un "1" y a continuación un "0". (No necesitaremos un estado nuevo para recordar el último elemento de la cadena.)

De forma similar procederemos en la impresión. Identificado el símbolo a imprimir en *M* procederemos en *M'* a imprimir la traducción de éste. Primero el símbolo que en ese momento escudriña *M'*, que pasará a un estado $(\eta\delta)_{01}{}^{D}$ en el que el superíndice (D o I) recordará el movimiento a izquierda o derecha que

debería hacer *M*. Como subíndice el resto de la cadena que queda por imprimir en orden inverso (el primer elemento de la derecha no se incorpora como elemento a recordar, pues se imprimirá en ese movimiento). Con "η" recordamos el estado a que debemos volver cuando se imprima la cadena tras pasar por los estados $(\eta\delta)_{01}{}^{D}$ y $(\eta\delta)_{0}{}^{D}$, en el ejemplo.

Así,

$$(\eta\delta)_{01}{}^{D}$$

quiere decir que resta por imprimir la cadena "01" y después producir un desplazamiento de *z* casillas a la derecha para escudriñar la correspondiente celda. (Recordemos que "z" fue definido al comienzo.)

Así, pues, en *M'* deberemos contar con todos los estados η de *M*. Para cada estado η un conjunto de estados η_0, η_1, η_{00} η_{01} para cada una de las combinaciones que se pueden obtener con elementos de {0, 1} tomados de *i* en *i* con repetición para 0<i<z, combinaciones que pasan a ser subíndice en dichas expresiones. Para cada estado η un conjunto de estados η_B, η_{BB}, $\eta_{BBB...}$ cuyos subíndices contienen hasta z-1 ocurrencias de "B". (Estos estados se utilizarán en la lectura.) Además para cada estado η de *M* dispondremos de un conjunto de estados $(\eta\delta)_0{}^{D}$, $(\eta\delta)_1{}^{D}$, $(\eta\delta)_{00}{}^{D}$, $(\eta\delta)_{01}{}^{D}$, $(\eta\delta)_{10}{}^{D}$, $(\eta\delta)_{11}{}^{D}$ para cada una de las combinaciones que se pueden obtener con elementos de {0, 1} tomados de *i* en *i* con repetición para 0<i<z, combinaciones que pasan a ser subíndice en dichas expresiones. El superíndice recuerda el movimiento a la derecha. Otro conjunto semejante pero con superíndice "I". Un conjunto de z-1 estados $(\eta\delta)_B{}^{D}$, $(\eta\delta)_{BB}{}^{D}$, con hasta z-1 ocurrencias de "B" como subíndice y un conjunto similar de conjuntos, pero con superíndice "I". (Estos estados servirán para el proceso de impresión.)

Una vez establecido el alfabeto de *M'* así como el conjunto de sus estados estableceremos el conjunto de sus funciones. Si

$\gamma(\eta,\sigma) = (\eta',\sigma', m)$ es una función de *M,* donde m es I o D,

necesitaremos en *M'* un paquete de funciones que lleven a cabo la operación que podríamos definir de la siguiente manera

$$\gamma(\eta, <\alpha_0\alpha_1\alpha_{0z-1}>) = (\eta', <\alpha'_0\alpha'_1\alpha'_{0z-1}>, m),$$

donde $<\alpha_0\alpha_1\alpha_{0z-1}>$ es la traducción de σ y $<\alpha'_0\alpha'_1\alpha'_{0z-1}>$ es la traducción de σ'. Se establecen de la siguiente manera.

Determinación de las funciones de M'

Sea $\gamma(\eta,\sigma) = (\varepsilon, \sigma', m)$ (donde m es I o D) una función de M. Y sea $\alpha_0\ \alpha_1 \ldots \alpha_{z-1}$ la traducción de σ y $\alpha_0'\ \alpha_1' \ldots \alpha_{z-1}'$ la traducción de σ'. A nuestra función de *M* le corresponde en *M*' el siguiente conjunto de funciones:

(1) $\gamma(\eta,\alpha_0) = (\eta_{\alpha 0,}\, B, D)$.

(2) $\gamma(\eta_{\alpha 0 \alpha 1 \ldots \alpha j}, \alpha_{j+1}) = (\eta_{\alpha 0 \alpha 1 \ldots \alpha j+1)}, B, D)$, para $j<z-1$, para todo estado η con subíndice formado por los j primeros elementos $\alpha_0\, \alpha_1 \ldots \alpha_j$ de la traducción de los símbolos de *M*.

(3) $\gamma(\eta_{\alpha 0 \alpha 1 \ldots \alpha z-2}, \alpha_{z-1}) = ((\eta\delta)^{m}{}_{\alpha'0 \ldots \alpha'z-2,} \alpha'_{z-1}, I)$, donde m es "I" o "D", según la función dada inicialmente.

(Con estas funciones se ha identificado la traducción del símbolo que escudriñaría *M,* y se imprime el último de sus símbolos. Con "m" se recuerda el movimiento que deberá hacerse cuado se termine la impresión.)

(4) $\gamma((\eta\delta)^{m}{}_{\alpha'0 \ldots \alpha'j,} B) = ((\eta\delta)^{m}{}_{\alpha'0 \ldots \alpha'j-1}, \alpha'_j, I)$ (donde m es "I" o "D"), para todo estado η con superíndice m y subíndice formado por los j primeros elementos $\alpha'_0\, \alpha'_1 \ldots \alpha'_j$ de la traducción de los símbolos de *M*.

(5) $\gamma((\eta\delta)^{m}{}_{\alpha'0,} B) = (\eta, \alpha'_0, m)$, donde m es "I" o "D".

(Concluida la impresión se vuelve al estado original η y se produce el movimiento "m" que corresponda.)

(6) Para todo i, tal que $1<i<z-1$, las funciones

$\gamma(\eta^{m}{}_i, 0) = (\eta^{m}{}_{i-1}, 0, m)$

$\gamma(\eta^{m}{}_i, 1) = (\eta^{m}{}_{i-1}, 1, m)$

$\gamma(\eta^{m}{}_i, B) = (\eta^{m}{}_{i-1}, B, m)$.

(7) Así como también

$\gamma(\eta^{m}{}_i, 0) = (\eta, 0, m)$

$\gamma(\eta^{m}{}_i, 1) = (\eta, 1, m)$

$\gamma(\eta^{m}{}_i, B) = (\eta, B, m)$.

En esta lista de funciones puede eliminarse las redundancias.

Esta máquina *M'* no está en condiciones de ser tratada por la máquina Universal de Turing presentada por Hopcroft-Ullman. La complejidad de los "nombres" de sus estados sólo responde al control del diseño de la transformación de la máquina dada.

Pero si ordenamos dichos estados con la única condición de que el primero ha de ser el estado inicial podemos identificar los estados con su número de orden en notación monaria. Esta notación pasará a ser su nuevo nombre: 1, 11, 111… etc. Si en las funciones anteriores introducimos estos nombres en sustitución de las complicadas denominaciones ofrecidas la máquina de Turing resultante está en condiciones de ser ofertada como input a la máquina universal de Hopcroft-Ullman.

La descripción de M' como input para la máquina universal de Turing

Para que la máquina universal pueda realizar el cómputo que haría la máquina *M* ante un determinado input θ aquella máquina debe recibir como input la "adecuada" descripción de la segunda máquina (en los términos que hemos presentado anteriormente: *M'*) así como la traducción del input sobre el que actúa *M.*

En la versión que se ofrece en Hopcroft-Ullman la máquina universal trabaja con una cinta dividida en dos pistas. Con ello resulta que dicha máquina trabaja con símbolos complejos, constituidos por un símbolo de la pista superior y su correspondiente pareja de la pista inferior. Esta circunstancia carece de importancia para nuestro problema. La pista superior, con sus símbolos, sólo son elementos de control de la máquina universal y no afectan para nada a nuestra máquina *M'*.

La pista inferior contiene un campo acotado para introducir nuestra máquina *M'*. Dicho campo está determinado por un símbolo especial al comienzo del mismo y otro al final. *M'* ocupará en este campo la pista inferior. A continuación deberá introducirse la traducción del input, también en la pista inferior. Si *M'* cuenta con un número "q" de estados el citado campo se encuentra dividido en q sectores, debidamente diferenciados por símbolos ubicados en la pista superior para que la cabeza lectora se dirija al primer sector si la máquina debe alcanzar el estado 1, al segundo sector, si debe alcanzar el estado 11, al tercer sector si debe pasar al estado 111 y así sucesivamente. Cuál sea el número que se adjudica a cada estado es irrelevante siempre que se reserve el número 1 para el estado inicial. Cada uno de los espacios reservados a los estados estará a su vez dividido en tres zonas, cuyo inicio y fin se encuentran perfectamente controlados por símbolos en la pista superior de la cinta tanto en su comienzo como en su final: una zona para cada uno de los tres símbolos de *M'* ("1", "0" y "B"). Cada zona estará ocupada por la función que determina la acción de *M'* cuando se encuentra en el estado al que pertenece el sector y escudriña el símbolo al que corresponde la zona.

Si se advierte que en las funciones se han sustituido los estados por cadenas de unos "1" y los restantes elementos son unos "1", ceros "0" o blancos "B" se comprende fácilmente que la máquina universal realiza el cómputo de *M'* ante ese input. Y por ello, el cómputo de *M* con el input del que es traducción el input dado.

La siguiente figura nos muestra cómo puede ser la cinta de la máquina universal conteniendo *M'* y la codificación del input de *M.*

Campo para la máquina codificada M'

		*																										
$	$	$				$				$					$	$				$					$				$	$

……..Sector para primer estado….. ……..…Sector segundo estado……..

…….(continuación de la cinta)……...

				*																								
	$	$	$	1	1	0	0	1	1	1	0	0	B	B	B													

La codificación *M'* de *M* se introduce en el campo comprendido entre las dos cadenas de tres símbolos "$". Dicho campo se haya dividido en sectores (uno por estado) delimitados por secuencias de dos símbolos "$", Dentro de estos sectores se establecen tres zonas separadas por un símbolo "$". Estas zonas se reservan para introducir las funciones correspondientes al símbolo "0", al símbolo "1" y al símbolo "B" cuando uno de ellos es escudriñado por la máquina en estado correspondiente al sector. El primer asterisco "*" de la pista superior servirá a la máquina universal para controlar el estado en que se encuentra *M'*. Con el segundo asterisco, se controla el símbolo que se escudriña en cada momento. El primero puede desplazarse por el campo de *M'* sobre el inicio del sector de un estado. A su vez el segundo asterisco por el campo del input, que se encuentra inmediatamente a continuación de la segunda serie de tres símbolos "$".

Referencias bibliográficas

Beneyto, R. (1995), 'Binarización de máquinas: un procedimiento', *Quaderns de Filosofia i Ciencia,* 25-26, pp. 117-122.

Hermes, H. (1961), *Aufzählbarkeit, Entscheidbarkeit, Berechenbarkeit*, Berlin, Springer. Versión inglesa de B. Eckmann y B. L. van der Waerden: *Enumerability, Decidability, Computability,* Berlin, Springer, 1969.

Hoptcroft, J. E. y Ullman, J. D. (1969), *Formal Languages and their Relation to Automata,* Reading, Massachusetts, Addisson-Wesley.

— (2006), *Introduction to Authomata Theory, Languages and Computation.* Reading, Massachusetts, Addisson-Wesley, 3ª edición.

Turing, A. M. (1936-37), 'On Computable Numbers, with an Application to the Entscheidungsproblem', *Proceedings of the London Mathematical Society* (ser. 2) 42, pp. 230-265; corrections ibid., 43 (1937), pp. 544-546.

On some natural generalizations of weak Kleene logic*

Rafael Beneyto y José Martínez Fernández
Universitat de València / Universitat de Barcelona
rafael.beneyto@uv.es / jose.martinez@ub.edu

The strong and weak Kleene schemes of interpretation for first-order logic are some of the more useful and widely applied three-valued schemes of interpretation. These three-valued logics have found many applications in philosophy and computer science. The Kleene logics are defined with the following operators on the set of truth values $E_3 = \{0,1,2\}$ (we use the subindex *s* for the strong and *w* for the weak operators; negation is common to both logics):

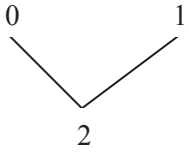
	$\neg$
0	1
1	0
2	2

$\wedge_s$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	2

$\wedge_w$	0	1	2
0	0	0	2
1	0	1	2
2	2	2	2

The rest of the propositional operators are defined in the usual way from negation and conjunction. In the philosophical applications of the theory, the value 1 corresponds to the value 'true' and the value 0 corresponds to 'false'. The value 2 has been applied to the sentences that are neither true nor false due to several semantic deficiencies: meaninglessness, paradoxicality and related forms of pathologicality (vicious circularity, ungroundedness), categorical mistakes, false presuppositions, undeterminateness (due to lack of reference, vagueness, etc.). If 2 means meaninglessness, the weak Kleene scheme is preferred, while if 2 represents undeterminateness the strong scheme seems more appropiate.

The Kleene logics are associated to a natural order of the truth values. All their operators are monotonic on the order of knowledge on E_3, the partial order (called E_3^k) determined by the following diagram:



This order reflects the degree of information about the truth value of a sentence. It is a natural requirement that operators should be monotonic on this order, since

* This work is being funded by the project HUM 2006-08236/FISO (C-CONSOLIDER) from the Spanish Ministerio de Educación y Ciencia.

increasing the information given by the arguments should not reduce the information given by the formula. Moreover, another natural condition on a k-valued operator in order to be a good generalization of a classical one is that it must be normal, that is, there must be two elements in the set of values such that they behave just like 'true' and 'false' do in classical logic. It is important to notice that Kleene negation is the only three-valued operator satisfying the conditions of normality and monotonicity with respect to the order of knowledge, and that the weak and strong Kleene conjunctions are the only two operators satisfying those two conditions plus commutativity.

A natural generalization of the strong Kleene scheme was given by Dunn and Belnap. Belnap's logic is defined by the following operators on the set of truth values $E_4 = \{0,1,2,3\}$:

	$\neg_b$
0	1
1	0
2	2
3	3

$\wedge_b$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	2	0
3	0	3	0	3

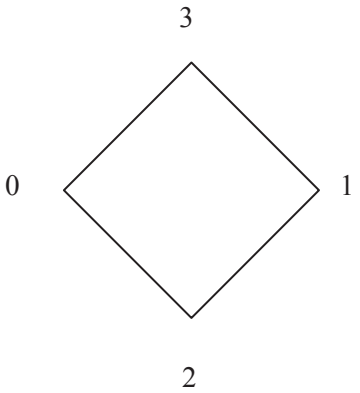
There are two main interpretations of Belnap's logic. The first one reads 0 as being assigned to false sentences, 1 the value for true sentences, 2 is the value assigned to neither true nor false sentences, and 3 to the sentences that are both true and false (the so called dialetheias).[1]

According to the second interpretation, that Melvin Fitting calls the "told" interpretation, each value corresponds to the set of truth values that a sentence is told to have.[2] Suppose we want to represent the information we get from some sources of evidence relative to the truth of a sentence. Four situations are possible: either all our informants say that the sentence is true, or all of them say it is false, or we receive no information, or we receive contradictory information: some of the experts say it is true, others say it is false. These different situations can be represented with the (epistemic) values 0, 1, 2, and 3, respectively.

The order of knowledge on E_3 can be extended to E_4. The order (which we will call E_4^k) has the following diagram:

[1] The concept was defined by Graham Priest. The existence of dialetheias is a hotly debated issue. It suffices for our purposes to say that some people defend the existence of dialetheias and they may want to have a use for this logic. (Priest himself argues that there are no sentences lacking classical truth value, and opts for a variation of Kleene's three-valued logic. However, his intuitions would justify Belnap's four-valued logic, if his arguments against the existence of sentences lacking truth value are wrong and there are such sentences. See Priest (2006), ch. 4,5.)

[2] Fitting (1994). This is coherent with the original motivation of Belnap, which was to give a logic to describe the semantic states a computer should have in order to classify the data it receives. See Belnap (1977).



Belnap's operators are normal and monotonic for E_4^k.

The aim of our paper is to determine several reasonable four-valued generalizations of the weak Kleene scheme, guided both by formal restrictions that the generalizations should satisfy and by the possible philosophical interpretations of the four values. All the operators should be normal, so that they can be seen as generalizations of the corresponding classical operator. Since Kleene operators are monotonic on the order of knowledge, a second natural condition is that the operators should be monotonic on the natural generalization of the order of knowledge on a set of four truth values. We will discuss whether the order E_4^k used by Belnap is philosophically cogent when trying to generalize the weak Kleene operators and we will argue that for some interpretation of the values 2 and 3 this is so, but for other interpretations other orders are preferable. The two conditions of normality and monotonicity for E_4^k determine uniquely an interpretation for negation: Belnap's negation. If we consider conjunction, we may add to normality and monotonicity the condition of commutativity, which is also natural.[3] These three conditions leave open nine possible conjunction operators. Adding the condition of associativity, only Belnap's conjunction and the following four operators are possible:[4]

$\wedge_{3w2s}$	0	1	2	3
0	0	0	0	3
1	0	1	2	3
2	0	2	2	3
3	3	3	3	3

$\wedge_{3w2w}$	0	1	2	3
0	0	0	2	3
1	0	1	2	3
2	2	2	2	3
3	3	3	3	3

[3] Commutativity is added just to simplify exposition: there are non-commutative variants of the following logics just as there are non-commutative variants of the Kleene operators and our results would apply also to those variations. They have applications in computer science to model the conjunction of computer programs operating in series.

[4] The logic with $\wedge_{3w2s}$ is analyzed in Fitting (1994) and the logic with $\wedge_{2w3s}$ is essentially the same. The other two logics, as far as we know, are new.

$\wedge_{2w3s}$	0	1	2	3
0	0	0	2	0
1	0	1	2	3
2	2	2	2	2
3	0	3	2	3

$\wedge_{2w3w}$	0	1	2	3
0	0	0	2	3
1	0	1	2	3
2	2	2	2	2
3	3	3	2	3

Notice that the operator $\wedge_{2w3w}$ can be obtained from the operator $\wedge_{3w2w}$ by permuting the values 2 and 3 in the truth table. Since this permutation leaves Belnap negation invariant, the logics generated by each operator are essentially the same. A similar connection obtains between $\wedge_{2w3s}$ and $\wedge_{3w2s}$. So we will restrict our attention to the operators $\wedge_{3w2s}$ and $\wedge_{3w2w}$.

In order to get good interpretations of these operators, it is important to see whether their restrictions to the sets {0,1,2} and {0,1,3} are a strong or a weak Kleene operator. It is also important to see whether some of the values are "contaminant" in the sense that any formula has that value if any of its arguments does.[5] Contaminant values can be assigned to sentences that are meaningless.[6]

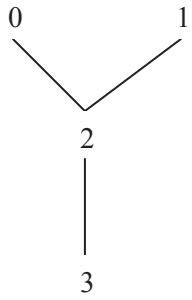
The dialetheist interpretation given for the Belnap operators could also be extended to these new operators, if one believes that sentences that are dialetheias behave like "contaminant formulas": every sentence in which a dialetheia occurs is bound to be a dialetheia. The option between $\wedge_{3w2s}$ and $\wedge_{3w2w}$ depends then on whether sentences that are neither true nor false behave in a weak or a strong way. In general, we think that these logics can be applied to analyze languages in which two different sources of semantic deficiency are present that deserve to be recorded separately. For instance, one could assign the value 3 to sentences that are meaningless, and assign the value 2 to sentences that lack truth value due to other deficiencies, like vagueness, names without referent, category mistakes, etc., having a strong attitude towards those sentences. Then the natural interpretation would be $\wedge_{3w2s}$. In an interpretation like this, sentences with value 2 are undetermined, they lack truth value but they are meaningful. On the contrary, sentences with value 3 lack truth value because they have no meaning, and they could not be true in any case.

Now a potential worry arises: in the dialetheist interpretation or in the epistemic interpretation of the truth values it makes perfect sense to use the order E_4^k as the order of information of the values, since having a sentence with value 3 (both true and false, complete evidence) gives more information about its truth status (or more evidence) than having values 1 (only true, only evidence in favour)

[5] This explains our name for the connectives: $\wedge_{2w3s}$ means that the restriction of the operator to the values {0,1,2} is the weak Kleene operator and that the restriction to the values {0,1,3} is the strong Kleene operator. The first number is always a contaminant value (this settles the value of $2 \wedge 3$ and $3 \wedge 2$ and completes the characterization of the truth table).

[6] Although not necessarily: in the strong Kleene scheme, the value 0, usually assigned to false sentences, is contaminant.

or 1 (only false, only evidence against), and these values give more information than having value 2 (lack of truth and falsity, no evidence). But in the interpretations that we are envisaging for the weak four-valued Kleene operators this order does not make sense anymore, since 3 is also assigned to sentences that are semantically defective, although in a more damaging way than the ones with value 2. The following order seems the sensible one to use, as it respects the order of defectiveness and also expands the order E_3^k:



It is a good feature of the operators $\wedge_{3w2s}$ and $\wedge_{3w2w}$ that they are monotonic for this order. The operators $\wedge_{2w3s}$ and $\wedge_{2w3w}$ are not monotonic, but this is not significant, since the relevant order for those operators is the one with 2 and 3 permuted (remember that in those operators it is the value 2 which is contaminant) and they are monotonic for this variant order. Notice that this order, while it is not a lattice, is still a ccpo (coherent complete partial order), so the logics still have interesting fixed-point properties (and they can be used in the study of self-referential languages, which are some of the main applications of this type of logics).[7] In an expansion of this paper we will explore some of the properties of these logics.

References

Belnap, N. (1977), 'A useful 4-valued logic', in Dunn, M. and Epstein, G. (ed.), *Modern Uses of Multiple-Valued Logics,* Dordrecht, Reidel, pp. 8-37.

Fitting, M. (1994), 'Kleene's Three Valued logics and Their Children', *Fundamenta Informaticae* 20, pp. 113-131.

Gupta, A. and Belnap, N. (1993), *The Revision Theory of Truth*, Cambridge, MIT Press.

Priest, G. (2006), *In Contradiction*, Oxford, Oxford University Press, 2nd edition.

Visser, A. (1984), 'Semantics and the Liar Paradox', in Gabbay, D. (ed.), *Handbook of Philosophical Logic*, Dordrecht, Reidel, vol. IV.

[7] See Visser (1984) and Gupta and Belnap (1993), ch. 2.

On the logical adequacy of identity criteria

Massimiliano Carrara and Silvia Gaio
University of Padua
massimiliano.carrara@unipd.it / silvia.gaio@unipd.it

Introduction: on the logical adequacy of identity criteria

In a realistic approach to ontology it is common to claim that we accept entities as real components of the world if and only if they belong to sorts for which identity criteria can be clearly stated. Identity criteria offer the conditions for determining when two individuals belonging to some sort K are identical.

Among the possible formulations of identity criteria, we consider the following:

(IC) $\forall x \forall y \in D(f(x) = f(y) \leftrightarrow R(x, y))$

It is assumed that there is a domain of individuals D and a function f such that $f(D)$ constitutes a sort of individuals K. R represents the condition under which x and y are said to be identical. In the left side of the biconditional in (IC), there is an identity relation, which is an equivalence relation. Consequently, the relation R on the right side of the biconditional must be an equivalence relation, too. Unfortunately, as has been observed in the philosophical debate about identity criteria, some relations considered as candidates for R often fail to be transitive. The following is an example of transitivity failure of R (see Williamson, 1986):

> Let x, y, and z range over colour samples and f be the function that maps colour samples to perceived colours. A plausible candidate for R might be the relation of perceptual indistinguishability. It is easy to verify, though, that such an R is not necessarily transitive: It might happen that x is indistinguishable in colour from y and y from z, but x and z can be perceived as different in colour.

The example shows how some relations that are intuitively plausible candidates to be identity conditions do not meet the logical constraint that (IC) demands. However, instead of refusing this kind of plausible but inadequate identity criteria, it has been suggested to approximate the relation R whenever it is not transitive. That means that, given a non-transitive R, we can obtain equivalence relations that approximate R by some operations.

Specifically, in this paper we analyze cases where the relation R which the identity condition consists of is not transitive and we seek for a way to obtain equivalence relations approximating R as much as possible.

Some approaches on how to approximate identity criteria have been suggested by Williamson (1986) and De Clercq and Horsten (2005). The aim of this paper is to present an improvement of De Clercq and Horsten's approach.

Closer approximations to identity conditions

Williamson (1986) suggests giving up the requirement for the identity condition to be both necessary and sufficient. Given a non-transitive R, let R_1, R_2, ... R_n be equivalence relations that approximate R. Among them, we want to find the relation R_i that best approximates R. Williamson's proposal is to apply one of the following approaches:

Approach from above: Consider the smallest (unique) equivalence relation R^+ such that $R \subseteq R^+$.

Approach from below: Consider the largest (not unique) equivalence relation R^- such that $R^- \subseteq R$.

Adopting the approach from above, you get a relation R^+ that is a sufficient identity condition. On the contrary, if you adopt the approach from below, you obtain a relation R^- that is a necessary identity condition. How can you choose between the two approaches? According to Williamson, there are non transitive relations R that are clearly necessary identity conditions (e.g., the relation of perceptual indistinguishability is considered a necessary identity condition for colours) and non transitive relations R that are clearly sufficient identity conditions (e.g., some forms of mental continuity are considered sufficient identity conditions for persons). So, when you deal with a non transitive, necessary identity condition, you apply the approach from below, otherwise you apply the approach from above.

De Clercq and Horsten claim that there are not always good reasons to decide whether you must take a necessary or a sufficient identity condition R. They consider a third option: to give up both the necessity and the sufficiency of the identity condition and to search for an overlapping relation $R^{\pm}$ that is neither a super- nor a sub-relation of R. Such an overlapping relation has the advantage of being closer to R than either R^+ or R^-. To understand why, consider how De Clercq and Horsten determine which relation R_i among the approximations R_1, ... R_n of a non transitive relation R is the closest (or best) approximation with respect to R. First, they call *revision* any adding or removing of an ordered pair to or from R; second, they count the number of revisions made to get each approximation R_1, ... R_n from R: such a number is called *degree of unfaithfulness*. Then, they state that a relation R_i is the best approximation with respect to R iff its degree of unfaithfulness is lower than that of all the other approximations of R.

Consider the following example. Let D be a domain of objects:

$$D = (a, b, c, d, e).$$

Assume there is a candidate relation R, reflexive and symmetric, for the identity condition for the individuals of D. Assume that for each element x of D, xRx. When R holds between two different objects x and y, we denote this as $\overline{xy}$. Let R on D be the following:

$$R = (\overline{ac}, \overline{ad}, \overline{bc}, \overline{bd}, \overline{cd}, \overline{de}).$$

R is not an equivalence relation. In fact, it fails to be transitive. For instance, R holds between a and d and between d and e, but it does not hold between a and e.

Apply, firstly, Williamson's approach from above. We obtain the smallest equivalence relation R^+ such that it is a superset of R, i.e.:

$$R^+ = (\ \overline{ab},\ \overline{ac},\ \overline{ad},\ \overline{ae},\ \overline{bc},\ \overline{bd},\ \overline{be},\ \overline{cd},\ \overline{ce},\ \overline{de}).$$

Apply, secondly, the approach from below. We get a relation R^- that is not unique. For instance, one of the largest equivalence relations that are subsets of R is the following:

$$R^- = (\ \overline{bc},\ \overline{bd},\ \overline{cd}).$$

Apply, then, the overlapping approach. You obtain the following relation:

$$R^\pm = (\ \overline{ab},\ \overline{ac},\ \overline{ad},\ \overline{bc},\ \overline{bd},\ \overline{cd}).$$

R^+ is obtained by adding four ordered pairs to R, R^- by removing three ordered pairs, and $R^\pm$ by adding one ordered pair and removing another one. The degree of unfaithfulness of R^+ is 4, the degree of R^- is 3, the degree of $R^\pm$ is 2. The latter is the lowest degree of unfaithfulness. Thus, $R^\pm$ is closer to R than R^+ and R^-. That means that with $R^\pm$, you stay closer to your intuitive identity condition R, because R^- modifies R less than R^+.

Refinement of the overlapping approach

If a relation R is not transitive and then, according to De Clercq and Horsten, possibly neither necessary nor sufficient, then it can occur that R holds between two objects a and b in a given situation, but in other situations R does not hold between the same objects a and b. We claim that such a variation of R with respect to a and b occurs both when we consider a and b in different contexts (i.e. in contexts containing a different number of objects) and when we consider a and b from different levels of observation. Consider the following two examples concerning perceived colours:

1. You see two monochromatic colour samples, A and B, and you do not see any difference with respect to their colour. Accepting that the identity condition for perceived colours is perceptual indistinguishability, you claim that A-colour is identical to B-colour. Now, add two further monochromatic colour samples, C and D, such that they are perceptually distinguishable. However, A is indistinguishable from C and B from D. In such a scenario, you can accept to revise your previous identity judgment and say that A-colour is not identical to B-colour.

2. You see two colour samples A and B from a distant point of view such that you are not able to distinguish A-colour from B-colour; so, you say that A-colour is identical to B-colour. Now you get closer to A and B and see a difference between them. So, you revise your previous judgment and say that A-colour is not identical to B-colour.

Our proposal is to integrate the notions of contexts and granular levels with De Clercq and Horsten's formal treatment of approximating relations. Informally, our

suggestion is as follows: Given a context, i.e. a set of elements of a domain, each granular level provides a relation R for the elements of that context; however, if we fix a granular level of observation, R can hold between two objects in a context and not hold between the same objects in a different context.

Let L be a formal language consisting of:

- individual constant symbols: $\overline{a}$, $\overline{b}$, … (there is a constant symbol for each element of the domain);
- individual variables: x_0, x_1, x_2, ... (countably many);
- two-place predicate symbols P_1, P_2, …; and
- usual logical connectives with identity, quantifiers.

The set of terms consists of individual constants and individual variable symbols. The formulas can be defined in the standard way.

Consider now the following interpretation of L. Let D_K be a fixed, non-empty domain of objects that we assume to belong to some sort K. A context o is defined as a subset of the domain D_K. So, the set of all contexts O in D_K is the powerset of D_K:

$$O = \wp(D_K).$$

Consider now a binary relation R (a two-place predicate). Assume that R is reflexive and symmetric, but not (always) transitive. R pairs the elements in each context $o \in O$ that are indistinguishable in some respect. For instance, in the case of color samples, R gives rise to a set of ordered pairs, each of them consisting of elements that are indistinguishable with regard to their (perceived) color.

Consider the behavior of R across granular levels. Take the following context with three elements: $o = (a, b, c)$. Depending on the granular level you are, one of the following scenarios can occur:

1. R gives rise to three ordered pairs.
2. R gives rise to two ordered pairs.
3. R gives rise to one ordered pair.
4. R does not give rise to any ordered pair.

In 1, we are in a coarse-grained level; in 4, in a very fine-grained level; and in 2 and 3, in some intermediate granular level. The extension of R for each context $o \in O$ varies across granular levels. Now, call *granular structure* a structure M consisting of the domain D_K and a binary relation R; formally, $M = < D_K, R>$. We assume that there is at least one granular structure for each granular level. Consider again the scenarios 1–4. There are very coarse granular structures with an R that behaves as in 1, some refined granular structures with an R that behaves as in 4, and other granular structures with an R that behaves as in 2 or 3.

Now, consider the behaviour of R across contexts. Given a granular structure, say M_1, consider two contexts: $o = (a, b, c)$, $o' = (a, b, c, d)$. Suppose that M_1 has a relation R such that $R_o = (\overline{ab}, \overline{bc})$ and $R_{o'} = (\overline{ab})$. You can observe that R holds

between *b* and *c* in *o*, but it does not hold between them in *o'*. So, fixed a granular structure, the extension of *R* can vary across contexts.

If, according to some granular structure, the relation *R* fails to be transitive with respect to some context $o \in O$, then the formal framework given by De Clercq and Horsten is applied. For instance, consider again M_1. Its relation *R* is not transitive in context *o*. Thus, an equivalence overlapping relation $R^{\pm}$ can be defined for *R* relatively to *o*. In contexts where *R* is not transitive, $R^{\pm}$ denotes a relation that differs from *R* in that it adds and/or removes some ordered pairs to or from *R*.

Conclusion

In this paper we have tried to show how the overlapping approach proposed by De Clercq and Horsten can be improved. Before determining the closest approximation to *R,* we suggest fixing a context and a granular level of observation, since *R* can vary along those two variables. If, according to a granular structure M_i, *R* fails to be transitive in a context, you can build the closest approximation to *R* for that context in M_i.

References

De Clercq, R. and Horsten, L. (2005), 'Closer', *Synthese* 146, n. 3, pp. 371-393.

Williamson, T. (1986), 'Criteria of Identity and the Axiom of Choice', *The Journal of Philosophy* 83, pp. 380-394.

Paraconsistent Vagueness: A Positive Argument

Pablo Cobreros
Universidad de Navarra
pcobreros@unav.es

Paraconsistent approaches have received little attention in the literature on vagueness (at least compared to other proposals). The reason seems to be that many philosophers have found the idea that a contradiction might be true (or that a sentence and its negation might both be true) hard to swallow. Even advocates of paraconsistency on vagueness seem to blush when they consider this fact; since they seem to have spent more time arguing that paraconsistent theories are at least as good as their *paracomplete* counterparts, than giving positive reasons to believe on a particular paraconsistent proposal (see, for example Hyde and Colyvan 2008). But it sometimes happens that the weakness of a theory turns out to be its mayor ally, and this is what (I claim) happens in a particular paraconsistent proposal known as subvaluationism. In order to make room for truth-value *gluts* subvaluationism needs to endorse a notion of logical consequence that is, in some sense, weaker than standard notions of consequence. But this *weakness* allows the subvaluationist theory to accommodate higher-order vagueness in a way that it is not available to other theories of vagueness (such as, for example, its paracomplete counterpart, *supervaluationism*).

Borderline-based theories and consequence relations

Most theories of vagueness take the notion of *borderline case* as a central one in the explanation of the phenomenon of vagueness. It is, therefore, natural for these theories to consider a notion of *definiteness* (represented as 'D') allowing us to talk about borderline cases. Several of these proposals share the common framework of possible-worlds semantics. Each theory differs on the informal reading of the semantics and, consequently, on the informal reading of D. For example, epistemicism reads possible worlds as a particular sort of *epistemic possibilities* and D expresses a particular form of *knowability* (the notion of borderline-ness expresses a form of unknowability). In turn, each reading of the semantics motivates a particular reading of truth, and since logical consequence is a matter of necessary preservation of truth, a particular notion of logical consequence. For example, for contextualism, that a sentence is true means that it is true in a particular context; this motivates the reading of logical consequence as preservation of truth at each context (this is local validity, '$\models_l$' below). Supervaluationism is usually associated to a stronger notion of consequence known as global validity ('$\models_g$' below). Subvaluationism is committed to a notion of consequence different to either local or global validity ('$\models_s$' below).

Definition 1 (Local consequence) A sentence ϕ is a local consequence of a set of sentences Γ, written $\Gamma \models_l \phi$, just in case for every interpretation and world *w* in the interpretation: if every member of Γ is true at *w* then ϕ is true at *w*.

Definition 2 (Global consequence) A sentence ϕ is a global consequence of a set of sentences Γ, written $\Gamma \models_g \phi$, just in case for every interpretation: if every member of Γ is true at every world then ϕ is true at every world.

Definition 3 (Subvaluationist consequence) A sentence ϕ is a subvaluationist consequence of a set of sentences Γ, written $\Gamma \models_s \phi$, just in case for every interpretation: if every member of Γ is true at some world then ϕ is true at some world.

Higher-order vagueness and consequence relations

In her 2003 paper Delia Graff Fara presents an argument against truth-value gap theories concerning higher-order vagueness. A theory of vagueness should explain at least why there seems to be no sharp transitions in sorites series. For example, in a sorites series for the predicate 'tall', there seems to be no sharp transition from the members of the series that are tall to those that are not tall. According to a truth-value gap theory this fact is explained by the existence of truth-value gaps: there seems to be no sharp transition from the members of the series that are truly tall to those that are truly not tall (that is, *falsely* tall), because there is at least a member in between such that it is neither truly tall nor truly not tall (that is, neither truly tall nor falsely tall). Thus, what explains the absence of sharp transitions in the series is the truth of the following principle (where 'D' is the object-language expression of the truth-gap theory's truth predicate):

(GP for 'tall') $\mathrm{Dtall}(x) \rightarrow \neg \mathrm{D} \neg \mathrm{tall}(x')$

(where x' is the successor of x)

One of the tough problems of vagueness is that the seeming absence of sharp transitions in the series does not stop here. There seems to be no sharp transition either between the truly tall's and the non truly tall's, nor between the truly truly tall's and the non truly truly tall's etc. In order to treat unrestricted higher-order vagueness at least as a logical possibility, the truth-gap theorist should be able to endorse the endless hierarchy of *gap principles* of this form:

(GP for 'D^ntall') $\mathrm{DD}^n\mathrm{tall}(x) \rightarrow \neg \mathrm{D} \neg \mathrm{D}^n\mathrm{tall}(x')$

The most well known truth-value gap theory is supervaluationism which is supposed to be committed to global validity but Fara (2003) shows that gap-principles are globally inconsistent for finite sorites series.

Fara's objection is intended against supervaluationism, but the argument affects the other *borderline based* theories as well. The reason is that gap-principles represent in an abstract form the general strategy of this sort of theories in order to explain the seeming absence of sharp transitions in sorites series. For example, for the epistemicist there seems to be no sharp transition from the tall to

the non-tall members in the series because the *clearly* tall members and the *clearly* not tall are separated by members that are not *clearly* tall and not *clearly* not tall. In a similar way, there seems to be no sharp transition from the *clearly* tall to the not *clearly* tall because there are members in the series that are not *clearly clearly* tall and not *clearly* not *clearly* tall... That is, gap-principles are compelling for each borderline-based theory, reading 'D' in each theory's preferred way.

In the case of theories committed to local validity, we might find models showing the local consistency of gap-principles. But we might find ways to adapt the argument for the local notion of consequence.

> **Definition 4** (Absolute definiteness) A sentence ϕ is absolutely definite at a world *w* just in case $\{D^n\phi \mid n \text{ in } \omega\}$ holds at *w*.

The idea is that ϕ is absolutely definite at *w* just in case ϕ takes value 1 at w, and so does $D\phi$, and $DD\phi$ etc. Making use of this notion of absolute definiteness, the following connection between global and local validity holds:

> **Claim 1** If $\Gamma \vDash_g \phi$ then $\{D^n\gamma \mid \gamma \text{ in } \Gamma, n \text{ in } \omega\} \vDash_l \phi$

That is, ϕ is a global consequence of Γ just in case ϕ is a local consequence of the *absolute definitization* of the elements in Γ. The connection shows that, although one can endorse gap-principles given local validity, one cannot endorse that these are *absolutely definite*. How bad is this last result? Surely, it is not as bad as lacking even the possibility of accepting gap principles plainly (as it happens in the case of global validity). But it seems to me that the result is bad enough. As pointed out before, gap principles describe a theory's solution to the question of the seeming absence of sharp transitions in sorites series (reading 'D' in the preferred way of the theory). If one cannot endorse that these principles are absolutely definite, then one seems to be committed to the idea that the theory itself is not absolutely definite (reading 'D' in the particular way of the theory). For example, for the epistemicist, it wouldn't be absolutely knowable whether the theory is right (and the theory going wrong should be always an epistemic possibility). For the contextualist reading, more dramatically I think, there must be contexts where some gap-principles are false, and thus, there are contexts where the theory itself goes wrong.

At this point, the weakness of subvaluationist's logic turns out to be a great advantage. Claim 1 no longer works when we substitute '$\vDash_l$' for '$\vDash_s$'. Furthermore, it might be proved that the absolute definitization of gap-principles is consistent for finite sorites series given subvaluationist logic. According to subvaluationist consequence, that a set of sentences is satisfiable means that there is a structure such that each sentence is true in at least one world in the structure; Claim 1 shows that the absolute definitization of gap principles cannot be true at the same world *w*, but the absolute definitization of each gap principle might be true at different (and non connected) worlds in a single interpretation.

Conclusion

The capability of endorsing the absolute definitization of gap-principles looks like an appealing feature for any borderline-based theory of vagueness; but this capability is restricted to theories committed to a weak enough notion of logical consequence. If vagueness is to be explained in terms of borderline cases, the foregoing results constitute a good argument in favour of the subvaluationist theory of vagueness. Some philosophers will still find the commitment to parconsistency as something hard to swallow and will probably consider that the result speaks against the whole borderline-based approach to vagueness. These philosophers think that, as Williamson says, 'dialetheism is a fate worse than death' (2006, p 387). To these I find appropriate Priest's own response to Williamson: 'I haven't died yet, so I'm not in a position to judge' (Priest, 2007).

References

Beall, J. C. (ed.) (2003), *Liars and Heaps: New Essays on Paradox*, Oxford, Oxford University Press.

Fara, D. G. (2003). 'Gap principles, penumbral consequence and infinitely higher-order vagueness', in Beall, J. C. (ed.) (2003). Originally published under the name 'Delia Graff'.

Fine, K. (1975), 'Vagueness, truth and logic', *Synthese* 30, pp. 265-300.

Hyde, D. (1997), 'From heaps and gaps to heaps of gluts', *Mind* 106(424), pp. 641-660.

Hyde, D. and Colyvan, M. (2008), 'Paraconsistent vagueness: Why not?', *Australasian Journal of Logic* 6, pp. 107-121.

Keefe, R. (2000), *Theories of Vagueness*, Cambridge, Cambridge University Press.

Priest, G. (2007), 'Review of Absolute Generality', *Notre Dame Philosophical Reviews*, <http://ndpr.nd.edu(review.cfm?id=11144>.

Rayo, A. and Uzquiano, G. (eds.) (2006), *Absolute Generality*, Oxford, Oxford University Press.

Varzi, A. (2007), 'Supervaluationism and its logic', *Mind*, 116(463), pp. 633-676.

Williamson, T. (1994), *Vagueness*, London, Routledge.

— (2006), 'Absolute identity and absolute generality', in Rayo and Uzquiano (eds.) (2006).

Abducción y revisión de creencias

Hans van Ditmarsch y Ángel Nepomuceno
Universidad de Sevilla
hvd@us.es / nepomuce@us.es

Introducción

La importancia de la abducción ha sido reconocida por investigadores líderes en diversos campos del conocimiento. Como se señala en Aliseda (2006), para J. Hintikka (Lógica y Epistemología), la abducción es el problema fundamental de la epistemología contemporánea; se trata de una forma de razonamiento que ha de verificar los requisitos o tesis siguientes (Hintikka, 1998): inferencial, de acuerdo con la cual la abducción es, o forma parte de, un proceso de inferencias; 2) de objetivo: se ha lograr la generación de hipótesis, de un lado, y, de otro, la selección de la mejor de ellas para su posterior análisis; 3) de comprensión: en la investigación científica la abducción comprende todas las operaciones por las cuales se elaboran las teorías científicas; 4) de autonomía, que aboga por la especificidad de la abducción: ésta es una forma de inferencia distinta de, e irreductible a, la deducción así como a la inducción. Por otra parte, para Herbert Simon (Psicología Cognitiva), el proceso retroductivo (otro término para la abducción) es el tema central de la teoría de resolución de problemas en las Ciencias Cognitivas. Paul Thagard (Ciencias de la Computación) considera que diversos tipos de abducción juegan un papel fundamental como estrategias heurísticas en el programa IP, el cual, dicho esquemáticamente, es un sistema muy útil en representaciones y simulaciones computacionales de descubrimiento científico.

Organizamos el trabajo presentando en el próximo apartado un resumen de cómo se abordar la temática de la abducción desde un punto de vista lógico. Le sigue un apartado en el que nos referimos al modelo clásico *AGM* de revisión de creencias y sus operaciones epistémicas, la formulación de las mismas como operaciones de carácter abductivo, para introducir sucintamente el sistema axiomático básico *KD45*. Por último se presentan unas observaciones finales y una lista básica de referencias bibliográficas.

Estudio lógico de la abducción

Desde un punto de vista lógico, considerando que una teoría T es un conjunto de sentencias de un lenguaje dado, el específico de la teoría, y una observación o creencia expresada como una sentencia C de tal lenguaje, tras el proceso inferencial inherente a la abducción, se halla una nueva sentencia A como explicación, de manera que C, el *explanandum*, será una consecuencia de T y de A, que conjuntamente constituyen el *explanans*, en el sistema lógico (deductivo) de que se trate. Así pues, cabe presentar la abducción con estos tres elementos en el

siguiente esquema general: dado un problema abductivo (T, C), A es una solución si se verifica que

T, A implican C; es decir, $T, A \vdash C$

En definitiva, se han de tener en cuenta los siguientes elementos:

1. El inferencial, determinante de la relación que se establece entre *explanandum* (el hecho sorprendente C) y *explanans* (la teoría T junto con la solución hallada A). Esta relación podría ser la de consecuencia lógica clásica, como la indicada en el esquema, o una inferencia estadística o una relación de consecuencia no clásica. Dada la importancia histórica de los modelos deductivos de explicación, no nos ocupamos de formas de inferencia estadística y prestamos la mayor atención a las inferencias que han sido objeto de estudio de la lógica clásica y las que han constituido el objeto de las lógicas no clásicas.
2. Los *detonadores* abductivos. C puede ser un fenómeno novedoso, inesperado en cierto sentido, o bien se trata de una anomalía, un fenómeno que entra en conflicto con la teoría trasfondo T. En cada una de estas situaciones se plantean operaciones epistémicas distintas y relevantes para analizar sistemas de creencias.
3. Las soluciones. En el esquema citado, para el problema (T, C) A es una solución. No obstante, también es posible considerar problemas abductivos desde otro punto de vista, de manera que dado que $T \not\vdash C$ y $T \not\vdash \neg C$, se trate de obtener una nueva lógica (un conjunto de nuevas reglas de inferencia, por ejemplo) $\vdash'$ tal que $\vdash \subseteq \vdash'$, de manera que finalmente $T \vdash' C$. En ambas líneas son de aplicación las tablas semánticas, como, por ejemplo, las establecidas en Fitting y Mendelsohn (1998)

Ahora bien, a partir de la teoría T, no debemos olvidar que el hecho C tiene un carácter "sorprendente", se trata de una experiencia novedosa o bien se presenta como contraria a las expectativas. En concreto, se presentan dos posibilidades:

- C es una *novedad abductiva*, $T \not\vdash C$ y $T \not\vdash \neg C$: ni el hecho ni su negación tienen relación deductiva con la teoría
- C es una *anomalía abductiva*, es decir, de $T \not\vdash C$ pero $T \vdash \neg C$; en este caso la negación del hecho es deducible desde de la teoría

Revisión de creencias

En revisión de creencias *AGM* es el modelo clásico, cuyas operaciones epistémicas, a saber, *expansión*, *contracción* y *revisión*, pueden ser explicadas en términos abductivos. En concreto, las operaciones epistémicas que inducen cada uno de los detonadores abductivos son los siguientes

1. *Expansión abductiva*. C es una novedad abductiva, $T \not\vdash C$ y $T \not\vdash \neg C$; entonces se calcula A y la teoría se expande:

$$ExpAb(T, C) = Cn(T + A), \text{ con lo que } C \in ExpAb(T, C)$$

2. *Revisión abductiva*. Dada una anomalía abductiva, el proceso es el siguiente: se obtiene $T' = T - \{B_1,\ldots,B_n\}$, una de las posibles formas de *contracción*, de tal manera que $T' \not\vdash \neg C$, entonces hallar A tal que $T', A \vdash C$.

Por lo que respecta al modelo *AGM*, la operación de expansión que en el mismo se considera es equivalente a la indicada de expansión abductiva. En efecto, para cualquier base de conocimientos (que en el marco *AGM* viene a ser una teoría T, cerrada bajo operación de consecuencia), para cualquier sentencia B,

$$Exp(T, B) = Cn(T+B).$$

Dada la teoría base T, se comprueba que $ExpAb(T, C) = Cn(T+A)$, es decir, puesto que $ExpAb(T, C) = Cn(T+A)$ y $Exp(T, A) = Cn(T+A)$, $ExpAb(T, C) = Exp(T, A)$. En la contracción el problema es más complejo, como sucede con la revisión, considerando que en ésta intervienen tanto expansión como la propia contracción.

Si se toma el sistema de axiomas para creencias *KD45*, integrado por los que se indican más abajo, se presentan interesantes cuestiones al estudiar revisión de creencias y sistemas multi-agente. Más en concreto, se confrontarán las operaciones epistémicas desde un punto de vista abductivo, considerando el sistema *KD45*, el cual se define a partir de un lenguaje proposicional que contiene un operador de creencia "B" (en su caso alternativamente, un operador de conocimiento "K", aunque aquí sólo consideramos el anterior) y un determinado conjunto de agentes $\{a, b,\ldots\}$; la semántica es la de tipo kripkeano y se consideran relaciones de accesibilidad para cada agente. Sus (esquemas de) axiomas son los siguientes:

1. Todas las tautologías expresables en el correspondiente lenguaje proposicional
2. $B_a (A \rightarrow C) \rightarrow (B_aA \rightarrow B_aC)$; distribución de B_a respecto de la implicación
3. $\neg B_aA \perp$; creencias consistentes
4. $B_aA \rightarrow B_aB_aA$; introspección positiva
5. $\neg B_aA \rightarrow B_a\neg B_aA$; introspección negativa
6. La regla de *Modus ponens*
7. La regla de necesitación de creencias: de A se infiere B_aA.

Cabe una propuesta de extensión mediante abducción de teorías modales de creencias. A este respecto, para un conjunto de fórmulas modales de creencias contamos con dos maneras de tratar tal conjunto. Sea, por ejemplo, el conjunto $T = \{B_a(A \rightarrow C)\}$, si consideramos el cierre deductivo de T, *Cn(T)*, haciendo uso de las reglas de *KD45*, éste contiene las fórmulas

$$B_a B_a (A \to C),$$
$$(\neg A \wedge B_a (A \to C)) \vee (C \wedge B_a (A \to C)),$$
$$(A \to C) \wedge B_a (A \to C),$$
$$(\neg A \vee C) \wedge B_a (A \to C)),$$
etc.

En realidad el conjunto $\{B_a(A \to C)\}$ representa una creencia del agente a y si tal agente cree que $A \to C$, también cree que $B_a(A \to C)$, y así sucesivamente. Con esta interpretación se permite tanto la expansión como la revisión abductivas. Sea una teoría $T = \{B_a(A \to C)\}$, tal que se produce B_aC como novedad abductiva, entonces por definición tenemos que T |-/- B_aC y T |-/- $\neg B_aC$. Una expansión de T que permita la explicación (deducción) de B_aC no es más que la expansión de la teoría con B_aA; dado que la teoría es el conjunto de las fórmulas creídas, ello también se aplica a fórmulas sin operadores modales: si $A \in T$, aunque se trate de una fórmula de esta clase, $B_aA \in T$.

Otra forma la clausura deductiva de una tal teoría T, que contiene sólo fórmulas creídas y no contiene, por tanto, las fórmulas no creídas, se plantea de la siguiente manera. Sea C una fórmula no creída, es decir, $C \notin T$; en tal caso, como el agente a no cree C, la teoría T debe contener $\neg B_aC$, de lo que se sigue que también contiene $B_a \neg B_aC$, y así sucesivamente. En ello consiste la propuesta de Segerberg (1999) para relacionar "revisión *AGM*" con lógicas dinámicas epistémicas. En esta perspectiva la expansión abductiva es imposible, sólo cabe la revisión abductiva, dado que un crecimiento de teorías que son deductivamente cerradas es imposible. Así, por ejemplo, para añadir B_aC a una teoría como la indicada que contenga $B_a(A \to C)$, e incluir todas las ignorancias, primero tiene que descartar $\neg B_aA$, entonces es necesaria una contracción previa a la expansión deseada con B_aA. Problemas similares se tratan en van Ditmarsch et al (2005).

Un desarrollo interesante sería permitir operadores dinámicos formalizando el resultado de expansión de la teoría T misma e investigar abducción a partir de estas condiciones. Por ejemplo, suponemos que $[*A]C$ formaliza que tras la expansión (o, en su caso, revisión) con A, se da C y que $T = \{[*A]C\}$; en este caso, si se elige la expansión abductiva de manera que $ExpAb(T, C) = Exp(T, A)$, entonces $ExpAb(T, C)$ |-- A: la propia forma de la teoría ya facilita una explicación del hecho sorprendente C, el cual ya no sería tan sorprendente.

Referencias bibliográficas

Aliseda, A. (2006), *Abductive Reasoning. Logical Investigation into Discovery and Explanation*, Dordrecht, Springer.

Fitting, M. y Mendelsohn, R. L. (1998), *First-Order Modal Logic*, Dordrecht, Kluwer.

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

Segerberg, K. (1999), 'Two traditions in the logic of belief: bringing them together', en H. J. Ohlbach y U. Ryle (eds.), *Logic, Language and Reasoning*, Dordrecht, Kluwer, pp. 135-147.

van Ditmarsch, H., van der Hoek, W. y Kooi, B. (2005), 'Public Announcements and Belief Expansion', en R. Schmidt et al. (eds.), *Advances in Modal Logic*, London, King's College Publications, pp. 335-346.

— (2008) *Dynamic Epistemic Logic*, Dordrecht, Springer.

La lógica dinámica topológica

David Fernández Duque
Universidad de Sevilla
dfduque@us.es

Lógica modal y topología

Históricamente, la semántica topológica para lógicas modales es anterior a la semántica de Kripke, siendo ésta conocida ya por Tarski y sus contemporáneos antes de 1940 (Tarski, 1938, pp. 103-134). Interpretamos fórmulas del lenguaje modal estándar sobre modelos topológicos, o bien un espacio topológico X dotado con una valuación V que asigna un subconjunto V(p) de X a cada variable proposicional.

Para extender V a una valuación de fórmulas arbitrarias, los operadores booleanos se interpretan de la manera acostumbrada, y el operador modal [] se interpreta como interior topológico; es decir, x satisface []A si en una vecindad de x, todo punto satisface A.

Esta semántica topológica genera la lógica modal S4, y dicha lógica es completa para interpretaciones en la recta real, en el espacio de Cantor y en topologías de Aleksandroff (básicamente, aquéllas que surgen de un marco de Kripke transitivo y reflexivo). Esto nos dice que S4 no es una herramienta demasiado útil si nuestro objetivo es clasificar espacios topológicos, pues es demasiado 'burda'.

Hoy en día se han dado varias propuestas para razonamiento espacial, en parte motivadas por aplicaciones en inteligencia artifical; (Aiello, 2007) muestra varias propuestas en el campo. Aquí nos interesan sistemas que agregan una componente temporal al S4 topológico.

Sistemas dinámicos topológicos

Una manera de representar el tiempo en un espacio topológico es a través de la acción de una función f. Un *sistema dinámico topológico* es una pareja ordenada <X,f>, en la cual X es un espacio topológico y f una función sobre X.

Esta definición general de un sistema dinámico nos permite estudiar estructuras provenientes de diversas ramas de las matemáticas de manera unificada. El ejemplo más típico de un sistema dinámico es un flujo en espacio euclideano; es decir, al resolver una ecuación diferencial obtenemos una función F(**x**,t), en la cual t es un número real y **x** un vector. Así tenemos que F(**x**,0)=**x**, ya que t=0 representa el estado inicial del sistema, y al variar el tiempo t los puntos se mueven a través del espacio.

Para acoplar el flujo F a nuestra definición de sistema dinámico topológico, basta tomar un valor fijo de t (digamos, t=1) y definir f(**x**)=F(**x**,1); si queremos lograr una mejor aproximación al sistema original, podemos tomar valores muy pequeños de t. Así, el flujo continuo avanza paso a paso en tiempo discreto.

Sin embargo, hay muchas otras clases de sistemas dinámicos, y no siempre provienen de un flujo en tiempo real; la teoría ergódica, por ejemplo, estudia sistemas dinámicos sobre espacios de probablidades. En este caso, además de la pareja <X,f>, contamos con una medida de probabilidad m, la cual satisface $m=mf^{-1}$.

Operadores temporales

En Artemov (1997) se define el sistema S4C, el cual agrega una segunda modalidad, O, a S4. Para interpretar fórmulas ahora necesitamos no un espacio topológico a secas sino un sistema dinámico topológico; seguimos interpretando a los operadores booleanos y [] como antes, además de definir

$$V(OA)=f^{-1}(V(A)).$$

Ahora podemos razonar no sólo sobre la estructura estática de X, sino también de lo que ocurre con éste al sufrir transformaciones. Este sistema aún no distingue marcos de Kripke de espacios topológicos en general, pero sí hay fórmulas que son válidas en la recta real pero no en otros espacios (Slavnov, 2003). Sin embargo, en Slavnov (2005) se muestra que S4C no distingue los espacios topológicos en general de los espacios euclideanos, aunque de dimensión indefinida; en Fernández Duque (2007) el autor demuestra que esto es cierto incluso si fijamos la dimensión igual a dos, es decir, cualquier fórmula satisfacible puede ser satisfecha en el plano euclideano.

Después, en Kremer (2005, pp. 133-158), se incorporó una nueva modalidad, *, a S4C; el sistema resultante es la *Lógica Dinámica Topológica*, abreviada DTL por sus siglas en inglés. El operador * se interpreta como un operador de preórbita bajo la acción de f; es decir, x satisface *A si todos los elementos del conjunto

$$\{x,f(x),f(f(x)),f(f(f(x))),...\}$$

satisfacen A. Estos sistemas ya distinguen a los espacios euclideanos de otros espacios topológicos (Fernández Duque, 2007).

Resultados conocidos

La sintaxis relativamente sencilla de DTL nos permite expresar de una forma compacta diversos fenómenos que pueden ocurrir dentro de dichos sistemas, así como ciertas propiedades propiamente topológicas. En este sentido nos sirve como una herramienta de clasificación. Algunas clases de sistemas que se pueden distinguir en DTL de sistemas arbitrarios son

- sistemas basados en espacios localmente conexos;
- sistemas basados espacios métricos completos;
- sistemas basados en espacios de probabilidades, donde la probabilidad es invariante bajo f;
- sistemas basados en marcos de Kripke.

El objetivo de nuestra investigación es describir qué clases de sistemas se pueden distinguir en DTL y cuáles no, así como dar axiomatizaciones u otros sistemas para demostrar los teoremas de DTL.

Más específicamente, si tenemos una clase K de sistemas dinámicos topológicos, ésta genera una lógica DTL(K) de todas las fórmulas de DTL que son válidas en K; es decir, todas las fórmulas A tal que para todo sistema dinámico <X,f> y toda interpretación V de las variables proposicionales, se cumple que V(A)=X.

Aquí tenemos resultados interesantes, ya que las propiedades computacionales de DTL(K) varían de manera muy drástica según K. Los resultados clave son

- si C es la clase de todos los sistemas dinámicos en los que f es continua (lo cual es la clase más general que consideraremos), entonces DTL(C) es indecidible (Konev, 2006, pp. 299-318), aunque en Fernández Duque (2009, pp. 110-121) el autor demostró que es recursivamente enumerable;
- si H es la clase de todos los sistemas dinámicos en los que f es un homeomorfismo, entonces DTL(H) no es recursivamente enumerable (Konev, 2004, pp. 182-196).

Direcciones futuras

El campo de la Lógica Dinámica Topológica es joven y aún quedan muchos problemas abiertos por resolver. Se tiene gran interés en comprender el comportamiento de sistemas basados en espacios compactos; sin embargo para esto no es suficiente utilizar la noción de validez, ya que las fórmulas válidas en espacios compactos son las mismas que en espacios arbitrarios. Una opción que puede rendir frutos es considerar alternativas 'existenciales' a la validez; por ejemplo, ¿podemos decidir si todo sistema dinámico compacto contiene un punto satisfaciendo una fórmula A dada?

Por otro lado, a pesar de que DTL(C) es recursivamente enumerable, esta enumeración no se da con un conjunto inteligible de axiomas, y el problema de axiomatizarla sigue abierto. Se ha propuesto una axiomatización, pero aún no se sabe si es o no es completa.

Referencias bibliográficas

Aiello, M., Pratt-Harman, I y Van Benthem, J. (2007), *Handbook of Spatial Logics*, Berlín, Springer.

Artemov, S. N.; Davoren, J. M. y Nerode, A. (1997), 'Modal Logics and Topological Semantics for Hybrid Systems', *Technical Report MSI 97-05,* Cornell University.

Fernández Duque, D. (2007), 'Dynamic Topological Completeness for R^2, *Logic Journal of IGPL,* doi: 10.1093/jigpal/jzl036.

— (2009), 'Non-deterministic semantics for dynamic topological logic', *Annals of Pure and Applied Logic* 157, n. 2-3, pp. 110-121.

Gabelaia, D., Kurucz, A., Wolter, F. y Zakharyaschev, M. (2006), 'Non-primitive recursive decidability of products of modal logics with expanding domains', *Annals of Pure and Applied Logic* 142, n. 1-3, pp. 245-268.

Konev, B., Kontchakov, R., Wolter, F. y Zakharyaschev, M. (2006), 'Dynamic topological logics over spaces with continuous functions', en Governatori, G., Hodkinson, I. y Venema, Y. (eds.), *Advances in Modal Logic* 6, pp. 299-318, London, College Publications.

— (2004), 'On Dynamic Topological and Metric Logics', *Proceedings of AiML 2004,* pp. 182-196.

Kremer, P. 'The Modal Logic of Continuous Functions on the Rational Numbers', manuscript.

Kremer, P. y Mints, G (2005), 'Dynamic Topological Logic', *Annals of Pure and Applied Logic* 131, pp. 133-158.

— (2007), 'A Chapter on Dynamic Topological Logic', en Aiello, M., Pratt-Harman, I. y van Benthem, J. (eds.), *Handbook of Spatial Logics*, Berlín, Springer.

Slavnov, S. (2005), 'On Completeness of Dynamic Topological Logic', *Moscow Mathematical Journal* 5, n. 2, pp. 477-492.

— (2003), 'Two Counterexamples in the Logic of Dynamic Topological Systems', *Technical Report TR-2003015,* Cornell University.

Tarski, A. (1938), 'Der Aussagenkalkül und die Topologie', *Fundamenta Mathematicae* 31, pp. 103-134.

Algunos criterios computacionales para los condicionales*

Santiago Fernández Lanza
Universidad Complutense de Madrid
flanza@filos.ucm.es

Introducción

La estrategia de formalización más común en Lógica es la de identificar una serie de expresiones lingüísticas como correspondientes a cada una de las expresiones lógicas que se consideren en el sistema formal con el que estemos trabajando. Para el caso de los condicionales, algunas de estas expresiones lingüísticas serían "Si A entonces B", "Si A, B", "Cuando A, B", "Sólo si B, A", etc. En todo caso, esta metodología de formalización no es infalible, sobre todo si se realiza atendiendo a cuestiones puramente sintácticas. Las expresiones por sí mismas no suelen permitir distinguir el tipo de condicional (material, causal, etc.), ni identificar el tipo de argumento (deductivo, inductivo, abductivo, etc.) en el que incorporar tales enunciados. Solemos utilizar las mismas palabras ("si … entonces …", "cuando …, …", etc.) para expresar condicionales tanto dentro de un argumento deductivo como en un argumento inductivo o abductivo, … Sin embargo, es posible poner en duda que el hablante que emplee tales expresiones esté queriendo decir lo mismo en cada caso o, dicho de otra forma, al oyente de tales proferencias le cabe esperar más o menos confirmación de la verdad del consecuente ante la confirmación de la verdad del antecedente. Por ejemplo, cuando ante un condicional como "Si estoy de pie y tengo las gafas puestas entonces tengo las gafas puestas" se confirma la verdad de su antecedente, la verdad del consecuente está necesariamente garantizada. Resulta imposible pensar en una situación que satisfaga el antecedente y no satisfaga el consecuente. Esta garantía no es tan fuerte en el condicional "Si es español entonces es de raza blanca" dado que la confirmación de la verdad del antecedente garantiza la verdad del consecuente no de forma necesaria sino probable. Se puede pensar perfectamente en una situación en la que el antecedente se satisfaga y el consecuente no se satisfaga y no por ello esto nos tiene que llevar al colapso o a reprocharle al hablante que lo que acaba de decir es falso porque hemos encontrado un contraejemplo. En tal situación, el hablante puede defenderse diciendo "yo estaba hablando en términos generales" que podría ser algo así como "yo no estaba deduciendo 'ser de raza blanca' a partir de 'ser español', sino que lo estaba induciendo debido a que la mayoría de los españoles son de raza blanca". A pesar de todo, estos condicionales son ampliamente utilizados y altamente operativos en muchas ocasiones. Se pueden plantear

* El presente trabajo ha sido financiado por los proyectos HUM2006-04955/FISO (Ministerio de Educación y Ciencia), FFI2008-03902 y FFI2009-08828 (Ministerio de Ciencia e Innovación).

ejemplos donde es menor la fortaleza de la garantía de la confirmación de la verdad del consecuente respecto a la verdad del antecedente, como en el caso "Si estás comiendo chiles entonces se te ha curado la úlcera de estómago" donde la conexión entre antecedente y consecuente no es necesaria ni probable, sino meramente conjetural. Se puede pensar en una situación en la que se satisfaga el antecedente y no se satisfaga el consecuente y ni siquiera podemos responder a ella diciendo "estoy hablando en términos generales, ya que a la mayoría de los que comen chiles se les ha curado la úlcera de estómago".

Sin embargo, una de las ventajas de la mencionada estrategia de formalización es que, por su carácter puramente sintáctico, se puede tratar computacionalmente. En este trabajo se intentará indagar en la cuestión de si existen más criterios sintácticos además de las meras expresiones lingüísticas ("si … entonces …", "cuando …, …", etc.) que puedan proporcionar mayor información sobre el tipo de condicional o el tipo de argumento en el que se podría incardinar el enunciado condicional en cuestión.

El tratamiento computacional de todas estas cuestiones, que es el objetivo final de este trabajo, hace que no nos podamos distanciar de las propuestas meramente sintácticas. Consiguientemente, este trabajo no debe interpretarse como un intento de establecer reglas generales e infalibles respecto a las cuestiones semánticas y pragmáticas que se derivan de todo este tratamiento, sino como el establecimiento de ciertos criterios o indicios que poseen cierto carácter aplicable.

Diversos usos de los condicionales

La riqueza y complejidad del lenguaje natural permite 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.

El contexto en el que, de forma estándar, los humanos utilizamos los condicionales es 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:

> Si estoy de pie y tengo las gafas puestas entonces *necesariamente* tengo las gafas puestas.
> Si es español entonces *probablemente* es de raza blanca.
> 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:

> 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:

> Si se calienta agua a más de 100º entonces (*naturalmente*) hervirá.
> 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:

> 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 contrafácticos 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:

> 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 condicionales 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:

> 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:

> 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:

> 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, pp. 109-132) y su versión explícita podría expresarse de la siguiente forma:

> 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:

> Si vas a salir entonces *te aconsejo que* lleves paraguas.
> Si vas a llegar tarde entonces *te ordeno que* no vuelvas a casa.
> 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 palabra clave estándar.

El tiempo verbal

En este apartado, se analizarán los distintos usos de condicionales propuestos anteriormente para hacer una estimación de la tolerancia a los cambios de tiempo verbal en el antecedente y el consecuente. La hipótesis inicial es que, para algunos usos, no son posibles todas las combinaciones de tiempos verbales en antecedente y consecuente.

Previamente, una regla que opera de una manera 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"). 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 Real Academia Española (1973, pp. 520-522) y Gómez Torrego (1997, pp. 322-327).

Las reglas que se pueden extraer de estos 12 usos de condicionales podrían ser las siguientes:

- 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.
- 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.

Conclusiones

A parte de la mera identificación de partículas como "si", "entonces", "cuando", etc. parece que existen 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. Este criterio no siempre es efectivo ya que no siempre hacemos explícito el uso del condicional que estamos utilizando. Sin embargo, su automatización no tiene una excesiva complejidad.

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 son válidas todas las combinaciones en todos los usos. Ordenando los ejemplos de menor a mayor grado de tolerancia podríamos resumir el presente estudio con el siguiente cuadro:

Antec	Consec	nec	reg	enf	consej	ame	prom	pred	ley	prob	Conj	iro	ofre
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

Si computacionalmente somos capaces de detectar el tiempo verbal del antecedente y el consecuente de un condicional, el cuadro anterior 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.

Referencias bibliográficas

Austin, J. L. (1956), 'Ifs and Cans', *Proceedings of the British Academy* 42, pp. 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, pp. 109-110.

Gómez Torrego, L. (1997), *Gramática didáctica del español*, Madrid, Ediciones SM.

Real Academia Española (1973), *Esbozo de una nueva gramática de la lengua española*, Madrid, Espasa-Calpe.

Sistemas expertos y lógica jurídica

Emilio García Buendía
Universidad Complutense de Madrid
egarciabue@yahoo.es

Presentación teórica

Leibniz, en su obra *Elementa iuris naturalis* ya había señalado las similitudes existentes entre la lógica modal alética y la deóntica dándose cuenta de la posibilidad de aplicar todas las técnicas de cálculo de la primera a la segunda. No obstante, leyendo su obra y su planteamiento se descubre que el proyecto de Leibniz es mucho más audaz y genial pues lo que se encuentra formulado expresamente en su obra es la búsqueda de la formulación de una verdadera teoría jurídica con una estructura perfectamente clara y definida con la rigurosidad que proporciona la lógica.

Si se analiza cualquier cuerpo legal y se reflexiona sobre sus contenidos se puede observar que una gran parte de las normas jurídicas se encuentran formadas por reglas. Pero estas reglas no contienen en modo alguno predicados deónticos en el sentido de *permitido, prohibido*, etc. sino que se las puede considerar como verdaderas reglas de producción.

De este modo, si un sujeto satisface dicho conjunto de reglas el ordenamiento jurídico le asigna un derecho subjetivo, lo modifica, lo extingue, transmite, etc..

Pero antes de seguir adelante es preciso fijar el concepto de *regla de producción*. Se entiende por regla de producción un enunciado que tiene la estructura:

SI <antecedente> ENTONCES <consecuente>

El lugar del <antecedente> viene ocupado por cualquier hecho o circunstancia mientras que el lugar del <consecuente> viene ocupado por cualquier acción. Si bien la estructura es muy similar a la de un condicional no tiene en modo alguno su naturaleza.

Es evidente que, por el principio de composicionalidad, los elementos que pueden ocupar el lugar del <antecedente> o del <consecuente> pues llegar a ser extraordinariamente complejos.

La idea de reglas de producción se encuentra directamente vinculada con los denominados sistemas expertos basados en reglas los cuales encadenan conjuntos de reglas de producción realizando inferencias. Cualquiera de estos sistemas expertos se encuentran formados por dos componentes básicos: 1) el motor de reglas y 2) la base de conocimiento.

El denominado *motor de reglas* se puede considerar como la aplicación informática cuya función consiste en explorar y aplicar el conjunto de reglas de producción a cada caso concreto evaluando cada una de ellas. Lo importante a

resaltar es que el motor de reglas o de inferencia es completamente independiente del conjunto de reglas de producción lo cual dota al sistema de una gran flexibilidad dado que puede evaluar un conjunto muy grande de reglas en función del área específica que se desee.

El denominado *base de conocimiento* se puede definir como el conjunto de reglas de producción relacionadas entre sí relativas a un área de conocimiento. Las reglas de producción suministran además la fundamentación de cada paso deductivo por lo que garantizan la corrección del razonamiento.

Volviendo de nuevo al ámbito jurídico, en esta primera fase de desarrollo de una teoría jurídica estricta se trataría de mostrar el hecho de que una gran parte del ordenamiento jurídico se encuentra formado por normas susceptibles de ser transformadas en reglas de producción lo cual implica inmediatamente la posibilidad de ser manejadas por un sistema experto y, por ende, evaluado por un computador.

Un claro ejemplo de este tipo de normas se puede encontrar en el art. 138 del vigente Código Penal en su redacción de 10/1995, de 23 de noviembre que dice:

> El que matare a otro será castigado, como reo de homicidio, con la pena de prisión de diez a quince años.

Este precepto penal es fácilmente transformable en una regla de producción cuya forma sería:

> SI una persona mata a otra persona
> ENTONCES será castigado con la pena de prisión de diez a quince años.

Dado precisamente la similitud formal con las expresiones condicionales, esta regla se puede a su vez transformar en una fórmula bien formada propia de Lógica de Primer Orden (LPO) lo que permite aplicar todas las herramientas propias de la Lógica matemática. De este modo se puede comprobar que, en una primera fase, grandes zonas del ordenamiento jurídico son susceptibles de ser transformadas en conjuntos de reglas de producción que un único motor de inferencia podría manejar automáticamente.

Aplicación práctica

Hasta este momento se han presentado las características que definen a un denominado sistema experto. Avanzando en el planteamiento propuesto se trata ahora de mostrar la posible aplicación concreta a algún aspecto específico del derecho; para ello se ha elegido el tema de la nacionalidad.

Teniendo en cuenta que el positivismo jurídico es la base filosófica sobra la que se sustenta nuestro sistema jurídico, la fundamentación de cualquier derecho subjetivo remitirá a las fuentes jurídicas que lo regulen. Por este motivo, las fuentes que ordenan todo lo relativo a la nacionalidad, expuestas según la pirámide normativa, son las siguientes: 1º) Constitución española, arts. 11 y 149, 2º) Código civil en su Libro Primero, Título Primero, arts. 17 y ss, 3º) Ley de 8 de junio de 1957 reguladora del Registro Civil, 4º) Decreto de 14 de noviembre de 1958 por el

que se aprueba el Reglamento de la Ley del Registro Civil, y 5º) Instrucciones de la Dirección General de los Registros y Notariado[1].

En principio, con este conjunto de normas se encuentra delimitado un verdadero conjunto el cual es susceptible de ser analizado con distintas herramientas lógicas a fin de comprobar su coherencia interna, analizar los elementos que los componen, las relaciones que existen entre ellos, etc. [2].

Si en las ciencias experimentales el problema de la fundamentación de la verdad de las proposiciones generales que conforman cualquier ciencia de este tipo sigue siendo uno de los problemas centrales de la filosofía de la ciencia, en una ciencia jurídica la cuestión de la fundamentación se resuelve por remisión a la norma legalmente aprobaba y promulgada; dicho en otras palabras, a través del conjunto normativo ya mencionado anteriormente.

El artículo concreto de la norma proporcionará la justificación de la modificación de cualquier derecho subjetivo de manera análoga a una regla de cualquier cálculo que justifica el paso de una deducción a otro garantizando la transmisión de la verdad.

Esta idea se puede ahora generalizar. De este modo, cualquier ámbito de la vida regulada por el derecho y dentro de un estricto positivismo jurídico, la adquisición, conservación, transmisión o extinción de cualquier derecho subjetivo encontrarán su fundamentación en un conjunto X de normas muchas de las cuales se pueden reescribir en forma de reglas de producción como ahora se verá.

Analizando las normas de esta materia en el Código Civil, en su redacción dada por la Ley 18/1990 de 17 diciembre (BOE 18/12/1990), la adquisición de la nacionalidad española puede obtenerse por cuatro vías: 1ª).- por origen (art. 17 CC), por opción (art. 20 CC), por carta de naturaleza (art. 21.1 CC) y por residencia (arts. 21.2, 22 y 23 CC).

En cada una de las vías anteriores para acceder a la nacionalidad se va enumerando toda una casuística y requisitos que debe cumplir aquella persona a la que se le pueda atribuir la nacionalidad española. Focalizando nuestro ejemplo en una de estas cuatro posibilidades, se pueden estudiar, por medio de un análisis de

[1] Hay que señalar que en este punto se ha recogido la normativa general si bien existen otras también aplicables en materia de nacionalidad que afectan a su regulación como son el Real Decreto 453/2004, de 18 de marzo, sobre concesión de la nacionalidad española a las víctimas de los atentados terroristas del 11 de marzo de 2004 (B.O.E. 22-03-04) y la Ley 52/2007, de 26 de diciembre (B.O.E. 27-12-07) conocida como *Ley de la memoria histórica*. En estas dos normas también se recogen aspectos que afectan a la obtención de la nacionalidad española en varios supuestos.

[2] Aquí se está realizando esta presentación a los fines deseados que es el mostrar cómo un sistema normativo se puede transformar en un sistema de reglas de producción. Desde un punto de vista lógico, el tema de la nacionalidad sería una clase que incorporaría dentro de sus elementos no sólo la normativa que afectase a las personas físicas sino también a las jurídicas y, generalizando al máximo, a cualquier ente susceptible de ser aplicado el predicado *nacionalidad* como podrían ser aeronaves, buques, entidades supranacionales tanto de derecho privado como público, entidades sin forma jurídica, etc.

casos, la adquisición por origen y las distintas circunstancias que conducen a su obtención.

El art. 17 CC, en su redacción dada por la Ley 18/1990 de 17 diciembre regula la adquisición de la nacionalidad por origen; ahora vamos a intentar transformar este texto legal en su equivalente de normas de producción.

El art. 17 1.a) que dice: *Son españoles de origen: a) Los nacidos de padre o madre españoles* podría rescribirse de la siguiente manera:

> **SI** nacido Y padre O madre son españoles
> **ENTONCES** nacional español

Pero esta regla de producción también se podría rescribir en lenguaje lógico formal del siguiente modo:

$$\forall x \left[Nx \wedge \left(Px \vee Mx \right) \rightarrow Ex \right]$$

siendo: x = un ente cualquiera

N= propiedad de haber nacido

P = propiedad de tener padre español

M = propiedad de tener madre española

E = propiedad de tener la nacionalidad española

Indudablemente, desde el punto de vista lógico, aquel ente que satisfaga dicha expresión adquirirá la nacionalidad española basado en lo dispuesto en el art. 17 1.a).

Continuando con lo recogido en este artículo 17 se analizan los casos de adquisición de nacionalidad en virtud del *ius soli*, es decir, por el hecho de haber nacido en territorio español.

El art. 17 mencionado continúa diciendo:

1. Son españoles de origen:

a)

b) Los nacidos en España de padres extranjeros si, al menos, uno de ellos, hubiera nacido también en España. Se exceptúan los hijos de funcionario diplomático o consular acreditado en España.

Lo dispuesto en la letra anterior se transformaría en una regla de producción con la forma siguiente[3]:

> **SI** nacido en España Y padre no español nacido en España
> O madre no española nacida en España
> **ENTONCES** nacional español

[3] Se excluye en este caso la excepción prevista en esta letra con la finalidad de simplificar pues su inclusión no supone ninguna dificultad teórica.

Del mismo modo que el caso anterior, esta norma tiene su equivalencia en lenguaje de la Lógica de primer orden del modo siguiente:

$$\forall xyz\ [Nx \wedge ([\neg Ey \wedge Ny] \vee [\neg Ez \wedge Nz]) \rightarrow Ex]$$

siendo: x = un ente cualquiera solicitante de la nacionalidad

y = un ente que tiene la propiedad de ser padre de x

z = un ente que tiene la propiedad de ser madre de x

N= propiedad de haber nacido en España

E = propiedad de tener la nacionalidad española

Con todo lo expuesto se considera que se ha mostrado la posibilidad de reescribir grandes áreas de un ordenamiento jurídico en reglas de producción las cuales también se pueden transformar en expresiones bien formadas de Lógica de Primer Orden. De este modo, todas las herramientas propias de la lógica matemática son perfectamente aplicables al lenguaje normativo sin necesidad de acudir a otras lógicas. En este ejemplo se muestra cómo los predicados *prohibido* y *permitido* propios de lógicas deónticas son superfluos.

Un primer paso pues en la elaboración de una teoría jurídica rigurosa y exacta pasaría pues por la identificación de dichas áreas del ordenamiento legal que se componen únicamente por normas transformables en reglas de producción. El ejemplo expuesto sobre la nacionalidad es perfectamente aplicable a otras materias tales como vecindad civil, ejecución de sentencias penales, etc.

Referencias bibliográficas

González Poveda et al. (2009), *Código Civil*, Madrid, Colex, 17ª edición.

Durkin, J. (1994), *Expert Systems: Design and Development*, New York, Maxwell Macmillan.

Leibniz, G. (1669), *Specimina Juris*, Darmstadt, Preussichen Akademie, Otto Reichl Verlag, 1830.

Nilsson, N. J. (2000), *Inteligencia artificial: una nueva síntesis*, Madrid, McGraw-Hill Interamericana de España S.A.

Una solución a la paradoja lógica de los dos sobres

Héctor Hernández Ortiz
IIF-UNAM

La versión probabilista de la paradoja de los dos sobres ha sido extensamente discutida. Varios autores afirman haber dado una solución plausible. Sin embargo, en 1993 Raymond Smullyan planteó una versión puramente lógica de la paradoja (una que no tiene que ver con probabilidades). Hasta el momento, esta variante lógica sólo ha recibido una propuesta de solución debida a James Chase (2002). El objetivo de este trabajo es:

(1) señalar por qué la propuesta de Chase resulta insatisfactoria.

(2) proponer una solución alternativa que resulta más plausible y simple de acuerdo con los criterios de solución comúnmente aceptados.

Planteamiento del problema

Se tienen dos sobres indistinguibles externamente. Uno de los sobres contiene el doble de dinero que el otro, pero no se sabe cuál. Se te permite elegir uno de ellos y quedarte con el monto que contenga. Existe la opción de cambiar el sobre antes de abrirlo. Si se pretende obtener el máximo monto posible, ¿es conveniente cambiar de sobre?

La versión no-probabilista de la paradoja surge del hecho de que se pueden probar, mediante argumentos puramente lógicos, las siguientes dos conclusiones contradictorias (*C*) y (*D*):

(*C*) Si ganas, el monto que ganarás es mayor que el monto que perderás si pierdes.

(*D*) Si ganas, el monto que ganarás *no* es mayor que el monto que perderás si pierdes. El monto a ganar o perder es el mismo.

A continuación se presentan las pruebas de (*C*) y (*D*) dadas por Smullyan:

Prueba de (*C*).

(1) Sea n el número de dólares en el sobre que estás sosteniendo ahora.

Entonces

(2) el otro sobre tiene o $2n$ o $n/2$ dólares.

Entonces

(3) si ganas en el intercambio, ganarás n dólares,

pero

(4) si pierdes en el intercambio, perderás $n/2$ dólares.

Puesto que

> *n* es mayor que *n*/2,
>
> entonces
>
> (*C*) el monto que ganarás, si ganas – el cual es *n* – es mayor que el monto que perderás, si pierdes – el cual es *n*/2.

Prueba de (*D*).

> Sea *d* la diferencia entre los montos de los dos sobres, es decir, sea *d* el menor de los dos montos. Si ganas en el intercambio, ganarás *d* dólares, y si pierdes en el intercambio, perderás *d* dólares. Y así los montos son los mismos después de todo. (Smullyan, 1993, pp.189-192).

Ahora bien, un argumento por reducción al absurdo para rechazar la conclusión (*C*) de que gano *más* de lo que pierdo en el intercambio, es que si así fuera, entonces aplicando el mismo razonamiento al otro sobre, se concluye que también me conviene cambiarlo. Así que aceptar *(C)* lleva a concluir que conviene cambiar el sobre elegido y al mismo tiempo no conviene cambiarlo. De modo que hay algún paso erróneo en la prueba de (*C*).

Solución de Chase

Al hacer explícita la conclusión (*C*), ésta afirma que:

> Existen *x* y *y* tales que:
>
> C_1: si ganas, ganas $x.
>
> C_2: si pierdes, pierdes $y.
>
> C_3: x>y.

Los antecedentes de C_1 y C_2 no pueden ser ambos verdaderos, así que al menos uno es contrario a los hechos. Puesto que no sabemos cuál es, ambos condicionales han de ser considerados contrafácticos.

Por consiguiente, si en vez de hablar en general de las cantidades *x*, *y*, representamos con *n* el monto real que hay en el sobre elegido, entonces la conclusión (*C*) expresada en términos de los condicionales contrafácticos correspondientes equivale a (*B*):

> B_1: El mundo posible más cercano en el cual ganas en el intercambio es un mundo donde ganas $*n*
>
> B_2: El mundo posible más cercano en el cual pierdes en el intercambio es un mundo en donde pierdes $*n*/2, y
>
> B_3: $n > n/2$.

Como *n* es un número positivo, (B_3) es claramente verdadero. Así que el problema estriba en los enunciados (B_1) y (B_2). Supongamos que el sobre elegido contiene el monto máximo, entonces (B_2) es verdadero porque el mundo posible más cercano donde pierdo es el actual y en él pierdo *n*/2. Pero como *n* es el monto máximo ¿qué valor de verdad tendrá (B_1)? Lo primero que hay que determinar es cuál es el mundo más cercano donde gano. Según Chase, ese mundo es M_1.

M_1: el mundo en donde se eligió el otro sobre.

M_2: el mundo donde se colocó en los sobres los montos n y $2n$ en vez de n y $n/2$.

Pero (B_1) *sólo* es verdadero si el mundo más cercano donde gano es M_2. Por consiguiente, (B_1) es falso. Análogamente, si el sobre elegido contiene el monto mínimo, entonces (B_1) será verdadero, pero (B_2) será falso. Ahora bien, en la prueba de (C) se infiere la verdad de *ambos* (B_1) y (B_2) a partir de la premisa verdadera (2): "Al cambiar el sobre ganarás $\$n$ o perderás $\$n/2$." Por lo tanto, el paso falaz en la prueba de (C) estriba en inferir la verdad *de los dos* enunciados contrafácticos (B_1) y (B_2) a partir de la verdad de la premisa (2).

Dificultades con la propuesta de Chase

Aquí se argumentará que la propuesta de Chase descansa sobre dos tesis cuestionables:

T_1: Los dos condicionales involucrados son contrafácticos.

T_2: En cada situación relevante uno de los dos condicionales es falso.

Contra T_1: Un problema con el argumento de Chase a favor de T_1 es que se puede concluir justo lo contrario usando un razonamiento análogo: "Puesto que en el intercambio ha de suceder que ganes o pierdas, entonces al menos uno de los dos antecedentes ha de estar de acuerdo con los hechos. Por lo tanto, es falso que ambos condicionales sean contrafácticos".

Contra T_2: El problema más serio en la propuesta de Chase es que lleva a otra paradoja. Si se acepta el argumento de Chase de que (C) es falsa porque incluye al menos un condicional falso, entonces la negación de (C) también es falsa. La razón es que, si los condicionales (B_1) y (B_2) son incompatibles, entonces (C) también sería falsa si en C_3 se sustituye '$x>y$' por '$y \geq x$', pero con esta sustitución se obtiene la negación de (C): "el monto que ganarás, si ganas *es menor o igual* que el monto que perderás si pierdes".

Por consiguiente, la propuesta de Chase conduce a la conclusión de que tanto (C) como su negación son falsas o, equivalentemente, que ambas son verdaderas por ser falsas sus respectivas negaciones. En particular, las condiciones que resultan en verificar un condicional y falsificar el otro en la prueba de (C) tienen el mismo efecto en los dos condicionales de la prueba de (D).[1] Puesto que los dos condicionales en (D) tienen el mismo antecedente que en (C), basta mostrar que sus consecuentes son equivalentes. Pero eso es claro porque "Ganas el monto actual" es equivalente a "Ganas la diferencia entre los dos montos" y "Pierdes la mitad del monto actual" es equivalente a "Pierdes la diferencia entre los dos montos". Así que si el argumento de Chase es correcto también se hace falsa la conclusión (D).

[1] De hecho, la simetría de las formulaciones entre (C) y (D) ha llevado a Albers (2005) a concluir que la paradoja es irresoluble.

Propuesta alternativa de solución

Esencialmente, la presente propuesta consiste en mostrar, mediante un examen riguroso de la prueba de (*C*), que la falacia involucrada es una falacia de ambigüedad, en particular, una falacia de equivocación. Se muestra que la conclusión (*C*) sólo se alcanza usando el mismo término con dos sentidos distintos en el razonamiento. Si el término equívoco se utiliza consistentemente en el argumento, la conclusión (*C*) no se sigue de las premisas.

Lo que mostraré a continuación es que, en la prueba dada, se utiliza un mismo término con dos sentidos distintos. El término equívoco es justo la variable *n*. En una parte de la prueba de (*C*) se usa para representar el monto mínimo, mientras que en otra representa el monto máximo.

Sabemos que

> (*) si ganas en el intercambio, entonces el sobre que estás sosteniendo ahora contiene el monto mínimo (de otra forma no podrías ganar al cambiar).

Ahora bien, de (1) y (*) se sigue que:

> Si ganas en el intercambio, *n* es el número de dólares en el sobre que estás sosteniendo ahora y el sobre que sostienes ahora contiene el monto mínimo.

Por lo tanto, en (3) *n* representa el monto mínimo. Análogamente,

> (**) si pierdes en el intercambio, entonces el sobre que estás sosteniendo ahora contiene el monto máximo.

Por lo tanto, por (1) y (**), la variable *n* en (4) representa el monto máximo.

Por lo tanto, la variable *n* se usa equívocamente en la prueba de (*C*).

Para confirmar que se trata efectivamente de una cuestión de uso equívoco de un término notaremos cómo un uso uniforme de la variable *n* evita el problema. Si representemos los montos en el sobre actual y en el otro sobre con *x*, *y* respectivamente en el par ordenado (*x, y*), entonces en la prueba de (*C*) se usan dos simbolizaciones: (*n*, 2*n*) y (*n*, *n*/2). Sin embargo, cómo se puede observar en la siguiente tabla, en cada simbolización el monto ganado es igual al monto perdido.

Situación	*Monto en el sobre actual*	*Monto en el otro sobre*
Ganas	*n*	2*n*
Monto ganado en el intercambio	*n*	
(Razonamiento a partir del otro sobre):		
Pierdes	2*n*	*n*
Monto ganado en el intercambio	*n*	

Situación	*Monto en el sobre actual*	*Monto en el otro sobre*
Pierdes	*n*	*n/2*
Monto perdido en el intercambio	*n/2*	
(Razonamiento a partir del otro sobre):		
Ganas	*n/2*	*n*
Monto ganado en el intercambio	*n/2*	

Obsérvese que si la variable *n* se usa consistentemente, las premisas no apoyan la conclusión (*C*), sino la (*D*), ya que en cada caso el monto a ganar o perder es el mismo.

Ahora bien, se podría objetar a favor de Chase que si se usa la misma simbolización coherentemente en todo el argumento al menos un condicional resulta falso. Por ejemplo, supongamos que usamos (*n*, 2*n*). En ese caso, la premisa (1) de la prueba de (*C*) es verdadera y el condicional (3) también. Entonces si pierdes, es porque elegiste el sobre con 2*n* y pierdes *n.* Pero entonces el condicional (4) es falso porque afirma que:

Si pierdes en el intercambio, perderás *n*/2.

Sin embargo, si aceptamos esta objeción, también tendríamos que aceptar que, contrario a la conclusión previa, la premisa (1) es falsa, ya que si tengo el sobre con 2*n*, entonces *n* no representa ya el monto actual, sino la mitad del monto actual. Además, la premisa (2) también resulta falsa porque al tener un sobre con 2*n*, no es cierto que el otro sobre tenga 2*n* o *n*/2. Esto sería un problema para Chase porque su propuesta requiere la verdad de la premisa (2). Sin embargo, como se muestra en la prueba de (*D*), un uso consistente de los términos no implica la falsedad de alguno de los dos condicionales, así que hay otra razón independiente para rechazar esta posible objeción.

En resumen, la paradoja lógica de los dos sobres surge por tener dos argumentos aparentemente sólidos a favor de conclusiones incompatibles. La propuesta de solución de Chase, entre otras dificultades, invalida ambos argumentos y hace falsas las dos conclusiones (*C*) y (*D*), lo cual lleva a una nueva paradoja. La presente propuesta ha señalado que sólo la prueba de (*C*) es falaz porque incurre en una falacia de equivocación. Si el término equivoco se usa consistentemente la conclusión que se sigue concuerda con la conclusión (*D*). De esta forma, la incompatibilidad desaparece y la paradoja se disuelve.

Referencias bibliográficas

Albers, C. (2005), 'Trying to resolve the two envelope problem', *Synthese* 145, pp. 89-109.

Chase, J. (2002), 'The non-probabilistic two envelope paradox', *Analysis* 62, pp. 157-160.

Clark, M. (2002), *Paradoxes from A to Z*, Londres y Nueva York, Routledge.

Katz, B. y Olin, D. (2007), 'A Tale of Two Envelopes', *Mind* 116, pp. 903-926.

Smullyan, R. (1993), *Satan, Cantor and Infinity: and other mind boggling puzzles*, Oxford, Oxford University Press.

Hacia una caracterización lógica del enfoque bayesiano de conocimiento común

Marco Antonio Hernández Ramírez
Universidad Nacional Autónoma de México
mahr@uxmcc2.iimas.unam.mx

En esta comunicación comparamos tres formalizaciones de la noción de conocimiento común presentes en la literatura: la de la lógica epistémica (Fagin et al., 1995, Mayer y van der Hook, 1995), la de la epistemología interactiva (Aumann, 1999) y la del enfoque bayesiano (Tan and Werlang, 1992). La definición lógica y la definición de Aumann son equivalentes. En este trabajo exploramos las relaciones entre esas definiciones y la caracterización de la noción bayesiana.

Panorama del conocimiento común

Intuitivamente, φ es conocimiento común para los agentes del conjunto *I* si todos ellos saben φ y todos ellos saben que *todos ellos saben* φ, y además todos ellos saben que *todos ellos saben* que *todos ellos saben* φ, y así sucesivamente.

El primer modelo formal del conocimiento es la lógica epistémica proposicional (Hintikka, 1962, Meyer y van der Hoek, 1995, Fagin et al., 1995) o de los mundos posibles. En este formalismo la noción de conocimiento es representada por un operador unario referido a agentes que precede las fórmulas de la lógica proposicional ***P***: $K_i\varphi$. La semántica de una fórmula epistémica se define por estructuras de Kripke, $\boldsymbol{M}=< S, \pi, R_1, ..., R_n>$, donde S es un conjunto no vacío de *estados*, π: $S\rightarrow (\boldsymbol{P}\rightarrow \{\mathbf{T},\mathbf{F}\})$ es una asignación de valores de verdad a fórmulas proposicionales atómicas por estado y $R_i \subseteq SXS$ (i=1,..., n) una relación de accesibilidad. La semántica kripkeana de una fórmula epistémica queda determinada por la relación $\models$, definida de la siguiente manera:

$$(\boldsymbol{M},s) \models K_i\varphi \rightarrow (\boldsymbol{M},t) \models \varphi \text{ para todo } t \text{ con } (s,t) \in R_i$$

Esta formalización nos permite representar de manera intuitiva propiedades del conocimiento tales como la introspección positiva o negativa ($K_ip \rightarrow K_iK_i\, p$ ó $\sim K_ip \rightarrow K_i \sim K_ip\}$). La referencia a agentes también nos invita a definir de manera natural conocimiento compartido por agentes, como el conocimiento mutuo ($E\varphi$, que leemos como "*todos saben φ*"), y conocimiento de orden superior, en particular, conocimiento común ($C\varphi$, que leemos como "*φ es conocimiento común*"). Informalmente:

$$E\varphi = K_1\, \varphi \wedge ... \wedge K_n\varphi$$

$$C\varphi = \varphi \wedge \mathrm{E}\varphi \wedge \mathrm{EE}\varphi \wedge ... = \wedge_{i\geq 0}\, E^i\, \varphi$$

La semántica de una fórmula con el operador de conocimiento común se define como sigue:

DEFINICIÓN 1: $(\boldsymbol{M},s) \models C\varphi \leftrightarrow (\boldsymbol{M},t) \models \varphi$ para todo t con s -» t.

La relación -» es la cerradura reflexiva-transitiva de la relación $\rightarrow$, donde $s \rightarrow_{Ri} t$ denota (s,t) $\in R_i$, con $s, t \in S$. Entonces s -» t se satisface si $s \rightarrow^k t$ para algún $k \geq 0$.

La primera referencia a la noción de conocimiento común aparece en Lewis (1969) en el contexto de la teoría filosófica de convenciones. Sin embargo, esta noción fue formalizada por Aumann (1976), en el marco de la teoría de juegos, de la siguiente manera. Sea (Ω, Σ, μ) un espacio de probabilidad donde: Σ es una Σ-álgebra sobre Ω, μ una medida de probabilidad sobre (Ω, Σ), $I=\{1,2, ..., n\}$ el conjunto de agentes y cada agente i es caracterizado por una sub-Σ-álgebra, $\prod_i$, que representa su información privada. Entonces:

DEFINICIÓN 2 (Aumann, 1976): Sea $\prod = \prod_1 \cap \prod_2 \cap ... \cap \prod_n$. Si el verdadero estado de la naturaleza es $\omega \in \Omega$. Entonces, un evento $E \in \Sigma$ se dice que es Conocimiento Común en ω si existe $B \in \prod$ con $\omega \in B$ y $B \subseteq E$.

La definición de conocimiento común canónica en el campo de la lógica se enmarca dentro de lo que Barwise (1985) consideró el enfoque iterado del conocimiento común. La definición de Aumann, por su parte, está en estrecha correspondencia con lo que podríamos considerar un enfoque de punto fijo, pero como él mismo afirma (1976, p. 1237), los dos enfoques representados por las definiciones anteriores, **Def. 1** y **Def. 2**, son equivalentes.

TEOREMA 1: Dada la estructura de Kripke $\boldsymbol{M}=<S, \pi, R_1, ... R_n>$ y $s \in S$, $\models_s C\varphi$ si y sólo si existe $B \in \prod$ con $s \in B$ y $B \subseteq \varphi$.

Críticas al modelo tradicional de conocimiento

La definición de conocimiento común propuesta por Aumann (1976, 1999) ha sido cuestionada en algunos aspectos, en particular, por ser definida como una iteración al infinito y por ser autorreferente (Tan y Werlang, 1992, Aliseda y Hernández, 2008). La definición de Tan y Werlang, con un enfoque bayesiano, da solución al problema de la autorreferencia presente en la definición de Aumann, aunque conserva su carácter de iteración al infinito.

Caracterización bayesiana del conocimiento común

El enfoque de conocimiento común de Tan y Werlang (1992) es bayesiano en espíritu y está construido sobre la estructura matemática de la recursión infinita de creencias. Para aspectos formales ver Brandenburger y Dekel (1993), Mertens y Zamir (1985), Harsanyi (1967-1968). Se asume con Aumann (1976, 1999) que el espacio básico de incertidumbre es Ω.

En el enfoque bayesiano cada agente asigna una probabilidad a la incertidumbre que encara. Entonces, sobre la incertidumbre de Ω, cada agente debe tener una creencia s_i^1 la cual es un elemento de

$$S_i^1 \equiv \Delta(\Omega)$$

Donde $\Delta(\Omega)$ es el conjunto de distribuciones de probabilidad sobre Ω. En adición a la incertidumbre del primer nivel, las probabilidades que los otros agentes

asignan a los elementos de Ω son también desconocidas para el agente *i*. Entonces, las probabilidades que los otros agentes asignan a Ω deben ser incluidas en la incertidumbre del agente *i*, lo cual conforma un segundo nivel de creencias donde el agente *i* asigna una probabilidad a las creencias de primer orden de los otros agentes y esta probabilidad (creencia) s_i^2 es un elemento de:

$$S_i^2 \equiv \Delta(\Omega \text{ X } \prod_{j\neq i} S_j^1)$$

Continuando con este proceso, se define inductivamente el m-ésimo nivel de creencias, donde el agente bayesiano *i* asigna una probabilidad a las creencias del (*m-1*)-ésimo orden de los otros agentes y esta probabilidad (creencia) s_i^m de m-ésimo orden es un elemento de

$$S_i^m \equiv \Delta(\Omega \text{ X } \prod_{j\neq i} S_j^{m-1})$$

La psicología o tipo [Harsanyi (1967)] del agente *i* está sintetizada por su jerarquía infinita de creencias:

$$s_i = (s_i^1, s_i^2,...) \in \prod_{j\geq i} S_j^m$$

El todo de la jerarquía infinita de creencias del agente bayesiano está contenido en un sistema de creencias [Aumann and Brandenburger (1995), p.1165]: dado un juego en forma estratégica, un conjunto $I=\{ 1, 2,...,n\}$ de jugadores, un conjunto de acciones A_i para cada jugador, un *sistema interactivo de creencias* para esta forma de juego consiste de un conjunto S_i, para cada jugador *i*, y para cada s_i de *i*:

- Una distribución de probabilidad sobre el conjunto S_{-i} de (*n-1*)-tuplas de tipos de los otros jugadores (la teoría de s_i)
- Una acción a_i de i (La acción de s_i) y
- Una función $g_i : A \rightarrow \boldsymbol{R}$ (la función de pago de s_i)

En este contexto, un estado es una descripción formal de las acciones de los agentes, sus funciones de pagos y sus creencias sobre las acciones y funciones de pago de cada uno de los otros agentes. En otras palabras, la teoría del tipo s_i representa las probabilidades que el agente *i* asigna a los tipos de los otros agentes.

Definición bayesiana de conocimiento común

Sea Ω, el espacio de incertidumbre y *E* un evento de interés. El agente *i* cree que *E* ocurre si y sólo si

$$s_i \in E_i^1 \equiv \{ s_i \in S_i \mid marg_\Omega [\Phi_i (s_i)](E)=1\}$$

La recursión infinita de creencias del agente *i* es tal que sus creencias sobre Ω asignan probabilidad uno al evento *E*. El agente *i* cree que todos creen que *E* ocurre si y sólo si

$$s_i \in E_i^2 \equiv \{ s_i \in E_i^1 \mid \forall k\neq i: s_k \in supp\ marg_{sk} [\Phi_i (s_i)] \rightarrow s_k \in E_k^1\}$$

Es decir, *i* cree que *E* ha ocurrido y si *i* cree que el agente *k* puede tener creencias s_k, entonces s_k debe pertenecer a E_k^1. Entonces el agente *i* cree que el agente *k* cree que *E* ha ocurrido.

El agente *i* cree que (*todos creen que*)$^{m-1}$ *E* ocurre si y sólo si

$$s_i \in E_i^m \equiv \{ s_i \in E_i^{m-1} \mid \forall k \neq i: s_k \in supp\ marg_{sk} [\ \Phi_i (s_i)] \rightarrow s_k \in E_k^{m-1} \}$$

La traducción es análoga sólo que ahora para el caso *m*-ésimo.

DEFINICIÓN 3 (Tan y Werlang, 1992): El agente *i* cree que *E* es conocimiento común, o a los ojos del agente *i*, *E* es conocimiento común si:

$$s_i \in \cap_{m \geq i} E_i^m \}$$

Donde $E_i^1 = E_i$ y para $m \geq 2$,

$$E_i^m = \{ s_i \in E_i^{m-1} \mid \forall j \neq i: supp\ marg_{sj} [\ \Phi_i (s_i)] \subseteq E_j^{m-1}$$

La recursión infinita de creencias ocurre en la mente de los agentes, de ahí el énfasis en que el evento es conocimiento común a los ojos del agente *i*.

TEOREMA 2 (Tan y Werlang, 1992, p. 160): Sea Ω finito y $\forall\ \omega \in \Omega$, $\mu(\omega) > 0$. Supongamos que $s_i \in P_i^{m-1}$ y que las creencias del agente *i* son consistentes con $\prod_i$ y μ en ω. En este caso $s_i \in A_i^m$, si y sólo si, $\omega \in K_i^m$.

El teorema **2** afirma que dada una estructura de Aumann, es decir, un espacio de probabilidad (Ω, Σ, μ) y una partición de información privada $\prod_i$, existe una estructura de recursión infinita tal que un evento *A* es conocimiento común, en el sentido de Aumann (**Def. 2**), si y sólo si *A* es conocimiento común en la recursión infinita (**Def. 3**).

Comentarios finales

La noción de Conocimiento Común in extenso es muy idealizada y no parece muy adecuada para modelar casos prácticos del mundo real. Buscamos una noción más realista, pero que conserve su carácter de conocimiento de orden superior, aunque no de iteración infinita (Aliseda y Hernández, 2008).

Nuestro punto de partida es encontrar la manera de representar en un lenguaje lógico la **Def. 3**. Esta definición salva un problema presente en la definición de Aumann (**Def. 2**), a saber, la autorreferencia. Entonces, la **Def. 3**, que es una formalización directa de la noción intuitiva de conocimiento común, es un buen candidato para formular una definición lógica de conocimiento común que no sea autorreferente.

Referencias bibliográficas

Aliseda, A. y Hernández, M. A. (2008) 'Lógica y Conocimiento: Notas sobre el Conocimiento Común', en *Memorias del XIII Congreso Nacional de Filosofía*, México.

Aumann, R. (1976) 'Agreeing to Disagree', *Annals of Statistics* 4, pp. 1236-1239.

— (1999) 'Interactive Epistemology I: Knowledge', *International Journal of Game Theory* 28(3), pp. 263-300.

Aumann, R. y Brandenburger, A. (1995), 'Epistemic Conditions for Nash Equilibrium', *Econometrica* 63, n. 5, pp. 1161-1180.

Barwise, J. (1987), 'Three Views of Common Knowledge', *Proceedings TARK-1987*, Los Altos, Morgan Kaufmann, 4, pp. 365-397.

Brandenburger, A. y Dekel, E. (1993), 'Hierarchies of Beliefs and Common Knowledge', *Journal of Economic Theory* 59, pp. 189-198.

Fagin, R., Halpern, J., Moses, Y. y Vardi, M. (1995), *Reasoning About Knowledge*, Cambridge, The MIT Press.

Harsanyi, J. C. (1967-1968), 'Games With Incomplete Information Played By 'Bayesian' Players, I-III', *Management Science* 14, n. 3, pp. 159-182.

Hintikka, J. (1962), *Knowledge and Belief*, Ithaca, NY, Cornell University Press.

Lewis, D. (1969), *Convention: a Philosophical Study*, Cambridge, MA, Harvard University Press.

Mertens, J. F. y Zamir, S. (1985), 'Formulation of Bayesian Analysis for Games With Incomplete Information', *International Journal of Game Theory* 14, n. 1, pp. 1-29.

Meyer, J. J. Ch. y van der Hoek, W. (1995), *Epistemic Logic for AI and Computer Science*, Cambridge, Cambridge University Press.

Tan, T. y Werlang, S. (1992), 'On Aumann's Notion of Common Knowledge: an Alternative Approach', *Revista Brasileira de Economia*, 46, n.2, pp. 151-166.

Complejidad y fundamentos de las ciencias deductivas*

Joost J. Joosten
Universidad de Sevilla
jjoosten@us.es

En varios campos de la ciencia —si no en todos— hay destacadas e importantes cuestiones sobre complejidad. Los ejemplos son abundantes: complejidad de problemas computacionales (P versus NP) en términos de cuánto tiempo necesita calcular un ordenador; complejidad de la solución de una ecuación diferencial, por ejemplo, en términos de periodicidad de la trayectoria calculada u otras propiedades topológicas; complejidad de teorías formales en términos de fuerza de la demostración; la complejidad de imágenes fractales en términos, por ejemplo, de dimensión fractal; la complejidad de lenguajes en términos de aprendizaje [posibilidad de aprendizaje]; la complejidad de un autómata celular en términos de un grado de aleatoriedad o en términos de poder computacional.

El programa propone estudiar interrelaciones entre varias clases de complejidad. ¿Qué tienen en común todas estas nociones de complejidad? Del mismo modo que Gödel aisló un conjunto de condiciones suficientes para la indecidibilidad, ¿es posible formular un conjunto (supuestamente) suficiente de condiciones para la complejidad intrínseca? Obviamente, la complejidad siempre es relativa a un marco cognitivo o formal. En cierto sentido, puede incluso entenderse como una consecuencia previsible al fijar los límites de este marco. El programa también tratará con esta cuestión.

Existe una fuerte convicción de que varios campos que conllevan complejidad pueden apoyarse mutuamente. En particular, nos interesamos por las relaciones entre lógica, computabilidad y lenguaje, por una parte, y por los sistemas dinámicos, fractales y autómatas celulares, por otra. Investigaciones llevadas a cabo recientemente ponen de relieve la existencia de profundas conexiones.

Como primer paso del programa, estudiamos los autómatas celulares en el sentido de Wolfram (2002, 1984, pp. 1-35), y en la ponencia presentamos diversos resultados sobre dichos autómatas. En primer lugar, consideramos autómatas que son unidimensionales en sentido de que constan de una cinta que se extiende en ambas direcciones hasta el infinito. La cinta está compuesta de células, cada una de las cuales puede tener un color. En el caso más simple, este color es blanco o negro. La cinta tiene una configuración inicial que asigna a cada célula un color.

* En esta ponencia el autor relata las líneas de su proyecto de Ramón y Cajal que entró en vigor en mayo del 2009 en la Universidad de Sevilla (Joosten, 2008). Entre otras cuestiones, el proyecto se centra en diferentes definiciones de complejidad en diversos campos de la ciencia relacionados con la lógica y en las interrelaciones entre estas definiciones.

Posteriormente la cinta se desarrolla en tiempo discreto, lo que significa que se evalúa paso a paso su nueva configuración. El color de una célula en el paso siguiente tan sólo depende del color de esta misma célula y de sus dos vecinos en el paso anterior, con lo cual sólo existen 256 distintos autómatas de este tipo. Sin embargo, exhiben una variada riqueza.

Ya en sistemas tan básicos como estos autómatas se puede observar comportamiento bastante complejo. Aun cuando estos sistemas poseen la complejidad Kolmogorov más baja posible, localmente pueden demostrar un comportamiento casi aleatorio. Y sin conocer la configuración inicial del sistema ni la regla que define el sistema, es en a veces imposible predecir el comportamiento local.

Podemos distinguir cinco tipos de complejidad en estos sistemas: regular progresivo, repetitivo, fractal, aleatorio, y aleatorio con pequeñas estructuras locales que interactúan de forma aleatoria. También se pueden aplicar otros tipos de complejidad a las autómatas celulares como la dimensión de Hausdorff. Estudiaremos cuestiones como si la dimensión de Hausdorff es sensible a la configuración inicial y cómo se relaciona con los cinco tipos de complejidad arriba mencionados. Estos resultados constituyen los primeros indicios de la posibilidad de relacionar entre sí diversos tipos de complejidad. Como epígono de este fenómeno mencionamos el teorema que afirma que, en un cierto sentido, las autómatas celulares mencionados arriba ya son Turing completos (Smith, 2007). La demostración de este teorema ha suscitado bastante polémica. En la ponencia detallaremos algunos de los argumentos más destacados de la misma.

Las máquinas de Turing se pueden considerar como autómatas celulares. La teoría nos dice que hay funciones que se pueden calcular con máquinas de Turing, pero el tiempo necesario para ello será intratable. Así que estas funciones, por muy complejo que sea el diseño de la máquina de Turing, necesitan mucho tiempo para ser computadas. En vez de trabajar con esas funciones dictadas por la teoría —que son más bien complejas y sólo implementables en máquinas de Turing con bastantes estados y con amplios símbolos para la cinta—, intentaremos buscar funciones con las mismas propiedades, pero que ya en máquinas de Turing muy sencillas destacan por su intratabilidad. Este trabajo relaciona la complejidad computacional con los modelos de computación y con la complejidad geométrica de las estructuras generadas por la máquina de Turing a lo largo del tiempo. Hay indicaciones de que es mucho más probable observar un fenómeno de 'slow-down' en vez de un fenómeno de 'speed-up', si incrementamos la complejidad de la máquina de Turing que podría computar una función. Una analogía de este fenómeno podría ser la razón clave para que un sistema natural evalúe para ser mas eficaz o no.

Referencias bibliográficas

Joosten, J. J. (2008), 'Complejidad y fundamentos de las ciencias deductivas', Solicitud de *Ramón y Cajal.*

Smith, A (2007), 'Universality of Wolfram's 2,3 Turing Machine', *The Wolfram 2,3 Turing Machine Research Prize.*

Wolfram, S. (1984), 'Universality and Complexity in Cellular Automata', *Physica 10D*, pp. 1-35.
Wolfram, S. (2002), *A New Kind of Science*, Champaign, Wolfram Media, Inc.

Los teoremas de completud y Leon Henkin

María Manzano
Universidad de Salamanca
mara@usal.es

Reconciliar las definiciones y los planteamientos sintácticos y semánticos de consecuencia constituye el núcleo de toda lógica —esto es, tanto de las lógicas puras y aplicadas, como de las clásicas y no clásicas—, incluso del de la que podemos denominar *lógica universal*.

El teorema de completud, junto con el de corrección, establecen la equivalencia entre la noción sintáctica y la semántica de consecuencia, para un cierto lenguaje. Podemos plantear la cuestión así: la noción semántica de verdad sirve para seleccionar del conjunto de todas las sentencias de un cierto lenguaje, a las que son verdaderas en todos las estructuras o modelos adecuados (las llamamos fórmulas lógicamente válidas, VAL). Por otra parte, a nuestro lenguaje formal, que es de naturaleza puramente sintáctica, podemos incorporarle un cálculo deductivo. Dicho cálculo permitirá deducir unas fórmulas de otras, y nos servirá para generar el conjunto de las sentencias del lenguaje que se pueden deducir sin premisas en el cálculo, a las que llamamos teoremas lógicos (TEO)

¿Coinciden esos conjuntos?

Demostrar que VAL$\subseteq$TEO es el objetivo del teorema de completud débil, que TEO$\subseteq$VAL lo es del de corrección. La demostración del teorema de corrección suele ser mucho más fácil que el de completud y es frecuente denominar teorema de completud a la igualdad entre los conjuntos VAL y TEO, esto es, a VAL=TEO.

La demostración de completud proporciona además información sobre la estructura de la clase de los modelos de una determinada lógica. Por ejemplo, La demostración de completud de Henkin para la teoría de tipos demuestra que el conjunto TEO de los teoremas del cálculo coincide con el conjunto VAL de las sentencias verdaderas en la clase de *estructuras generales*, que es un subconjunto propio del de las válidas en *estructuras estándar*. De hecho, cuanto más restrictiva sea la clase de modelos más amplia será la de las fórmulas en ella válida y a la inversa.

Relevancia teórica y práctica del teorema de completud

Importancia teórica: No sabemos qué es una lógica hasta que no hemos identificado al conjunto de sus fórmulas válidas; esto es *la logicidad* reside en ese conjunto. Me explico. Cada interpretación selecciona del conjunto de todas las fórmulas a las sentencias verdaderas en ella, que constituyen lo que denominamos *teoría de una estructura*, y que en principio será distinta para cada interpretación o modelo. Sin embargo, todas las teorías tienen un núcleo común, el de las fórmulas

válidas, VAL. Por ser verdaderas en toda estructura estas sentencias no describen a ninguna estructura en particular, sino a aquello que es común a todas ellas.

¿Caracterizan algo estas sentencias?

La respuesta es que sí, que *describen a la propia lógica*. Por consiguiente, si logramos generarlas con facilidad habremos *captado la esencia de la lógica*.

Importancia práctica: Aunque contemos con la noción semántica de verdad es frecuentemente muy difícil manejarla apelando simplemente a las condiciones de la definición. Mucho más difícil todavía lo es el determinar si una fórmula es consecuencia de un conjunto de ellas que tomamos como hipótesis; esto es, si toda interpretación en la que las hipótesis sean verdaderas la conclusión también. La razón es que en principio sería necesario comprobarlo en toda interpretación, modelo a modelo. Por fortuna hay otro modo de determinar si es una fórmula es consecuencia de otras que no es la mera verificación directa de sus especificaciones semánticas: se trata de inferir o deducir la fórmula en un *cálculo deductivo* utilizando las hipótesis; esto es, de establecer una cadena de razonamiento entre premisas y conclusión. De hecho, esta forma de definir el concepto de consecuencia es incluso más adecuada a la noción intuitiva, ya que refleja el carácter discursivo del razonamiento.

Si el cálculo deductivo nos va a resultar útil es porque nos ayudará a no equivocarnos; no nos conducirá de hipótesis verdaderas a conclusiones falsas: será un *cálculo correcto*. Además sus reglas permitirán obtener como teoremas a todas las consecuencias de un conjunto dado de hipótesis; esto es, será de aplicabilidad general, lo que denominamos un *cálculo completo.*

La historia de las demostraciones de completud

Hay tres ejes básicos en este estudio de la completud, que son:

1. **Origen:** ¿Cuándo y cómo nace la necesidad de demostración del teorema de completud?. ¿Cuándo se deslinda del teorema de decidibilidad? Nuestra hipótesis es que separar el concepto de completud del de decidibilidad no debió resultar fácil puesto que la primera prueba de completud se hizo para la lógica proposicional, que es completa y decidible; esto es, para ella existe un algoritmo que en un número finito de pasos te dice si la fórmula es válida o nó.
2. **Evolución de la prueba de completud de Henkin:** La demostración que hizo Henkin del teorema de completud resultó ser muy versátil y se pudo modificar para otras lógicas. ¿Cuál es la relación entre esa demostración y las pruebas de completud posteriores?. Nos parece especialmente interesante establecer la relación con las demostraciones que emplean *Conjuntos de Hintikka*, con las de naturaleza netamente algebraicas que emplean *álgebras booleanas, filtros* o *ultrafiltros*, las que construyen *modelos canónicos* para lógicas modales y temporales, o las que se valen de la introducción de *designadores rígidos* para lógicas híbridas.

3. **Pruebas alternativas de completud:** También en la historia de la lógica hay numerosos ejemplos de demostraciones indirectas de completud; por ejemplo, completud vía interpolación, completud vía compacidad, completud vía traducción a una lógica marco que sea completa, completud mediante modelos no estándar para lógicas incompletas. Será interesante incluir estas demostraciones en nuestra concepción general del concepto de completud para saber cómo y cuando debemos aplicar cada una de ellas.

Nociones ligadas a la de completud

Hay lógicas clásicas, proposicionales y de primer orden, también de orden superior. Hay una gran variedad dentro de la categoría de las lógicas denominadas no clásicas: abductivas, algebraicas, condicionales, combinatorias, categoriales, constructivas, cuánticas, deónticas, descriptivas, difusas, epistémicas, ecuacionales, estoicas, generales, libres, híbridas, infinitarias, intensionales, intuicionistas, lineales, multimodales, no monotónicas, paraconsistentes, polivalentes, de la relevancia, subestructurales y en general, una extensa clase de lógicas no-estándar. Claramente esta enumeración no es una buena clasificación ya que no es exhaustiva y es solapante (por ejemplo, hay lógicas modales que son intuicionistas y no-monotónicas). De hecho, la mayoría de los lógicos actuales piensan que la división entre lógica clásica y no clásica carece de sentido y sólo preserva una cierta connotación histórica.

Hemos seleccionado algunos sistemas lógicos cuyo estudio nos parece de interés:

1. Incluimos la lógica clásica, tanto la proposicional (LP), como la de primer orden (LPO) y la de orden superior (LOS). Su estudio no sólo es históricamente relevante, además nos proporcionan ejemplos de un comportamiento muy variado respecto al tema que nos ocupa :
   - para LP hay cálculos correctos y completos y normalmente terminan (esto es, proporcionan un procedimiento efectivo que permite listar todos los elementos del conjunto de teoremas lógicos, TEO);
   - hay cálculos correctos y completos para LPO, pero ninguno termina ya que el problema de la satisfacibilidad para la lógica de primer orden es indecidible.
   - LOS es incompleta con la semántica estándar, pero se pueden definir cálculos completos cuando las fórmulas se interpretan en *modelos generales*.
2. Mencionaremos algunos sistemas lógicos no clásicos porque la variedad de demostraciones del teorema de completud es muy rica e interesante:
   - Las lógicas multimodales normalmente emplean la construcción del modelo canónico para obtener resultados de completud de carácter general, tales como los teoremas de Sahlqvist.

- Las lógicas descriptivas proporcionan buenos ejemplos de resultados de completud (que terminan) para cálculos de tableaux usando la construcción de conjuntos máximamente consistentes.
- En lógicas híbridas normalmente se emplea la «construcción de Henkin» para obtener resultados de completud tales como el *Pure Formulas Theorem*.

3. También en la historia de la lógica hay numerosos ejemplos de demostraciones indirectas de completud; por ejemplo, completud vía interpolación, completud vía compacidad, completud vía traducción a una lógica marco que sea completa, completud mediante modelos no estándar para lógicas incompletas.

En la mayoría de los casos se cita la demostración de Henkin de 1950 o la de 1949.

Resultado universal de completud

¿Qué sentido tiene hablar de un resultado universal de completud?

Tenemos dos hipótesis de trabajo: La primera es que un resultado universal podría alcanzarse generalizando la idea de Henkin para elaborar su demostración de completud de primer orden a partir de su prueba de completud para teoría de tipos. De hecho pensamos que la demostración de completud de Blackburn para lógicas híbridas mediante designadores rígidos está en esa línea.

La segunda sospecha, aunque no contamos con demasiados indicios, es que una prueba de completud para lógicas que sólo poseen reglas estructurales está directamente ligado al tema de la invariancia. Tarski no se pregunta ¿qué es la lógica? ni tan siquiera ¿qué es una inferencia lógica? sino, ¿qué conceptos son lógicos? Usa su propia definición semántica de consecuencia, en la que sólo aparecen reglas estructurales, e intenta aplicar un programa que había dado grandes resultados en matemáticas: definir conceptos utilizando invariancias bajo ciertos grupos de transformaciones. De manera que un concepto es lógico si podemos definir transformaciones pertinentes en el universo que idealmente representa al de conjuntos del que extraemos las estructuras y demostrar la invariancia del concepto bajo las transformaciones. La jerarquía de tipos finitos es la idealización del universo más frecuentemente empleada y las transformaciones pueden ser permutaciones, biyecciones, isomorfismos, etc., según los distintos autores.

Referencias bibliográficas

Blackburn, P., de Rijke, M. y Venema, Y. (2001), *Modal Logic*, Cambridge, Cambridge University Press.

Blackburn, P., van Benthem, J. y Wolter, F. (eds). (2007), *Handbook of Modal Logic*, Amsterdam, Elsevier.

ten Cate, B. (2005), *Model theory for extended modal languages,* ILLC Dissertation Series DS-2005-01, University of Amsterdam.

van Benthem, J. F. A. K. (1979), 'Some Kinds of Modal Completeness', *Studia Logica* 39, pp. 125-141.

— (2009), *Modal Logic for Open Minds*, (en prensa).

Fine, K. (1978, 81), 'Model Theory for Modal Logic', Partes I-III. Part I y II en *Journal of Philosophical Logic* 7, pp. 125-156, 277-306; Parte III en *Journal of Philosophical Logic* 10, pp. 293-307.

Gabbay, D. y Guenthner, F. (eds.) (2001), *Handbook of Philosophical Logic*, Dordrecht, Kluwer Academic Publishers. 2ª edición.

Goldblatt, R. (1993), 'An Abstract Setting for Henkin's Proofs', en *Mathematics of Modality*, Standford, CA, CSLI Publications, pp. 191-212.

Henkin, L. (1950), 'Completeness in the theory of types', *Journal of Symbolic Logic* 15, pp. 81-91.

— (1963), 'En Extension of the Craig-Lyndon Interpolation Theorem', *Journal of Symbolic Logic* 28, n. 3, pp. 201-216.

— (1996), 'The discovery of my completeness proofs', *Bulletin of Symbolic Logic* 2, n. 2, pp. 127-58. (Presentado el 24 de Agosto de 1993 en el XIX International Congress of History of Science, Zaragoza, España).

Kripke, S. A. (1959), 'A completeness theorem in modal logic', *Journal of Symbolic Logic* 24, pp. 1-14.

Manzano, M. (1996), 'Extensions of first order logic. Number 19', en *Cambridge Tracts in Theoretical Computer Science*, Cambridge, Cambridge University Press, pp. 25-37.

— (1999), *Model Theory*. London, Oxford University Press.

Sahlqvist, H. (1975), 'Completeness and Correspondence in the First and Second Order Semantics for Modal Logic', en *Proceedings of the 3rd Scandinavian Logic Symposium*, Amsterdam, North-Holland, pp. 110-143.

Tarski, A. (1986), 'What are logical notions', *History and Philosophy of Logic* 7, pp. 143-212. (Texto de la conferencia de Tarski del mismo título, editado por Corcoran).

Logical Expressivism: Resnik's *versus* Field's*

Concha Martínez Vidal
University of Santigo de Compostela
mconcepcion.martinez@usc.es

Several problems in the philosophy of logic have lately received a lot of attention; among them, the problem of which logic is the right logic, the problem of its justification, and that of its normative character. My aim in this paper is to assess Field's "Logicism" in the light of Resnik's objection to it. With that purpose in mind, I analyse Resnik's and Field's positions on the abovementioned issues. The interest of comparing their views lies in the fact that though both are set forth by two empiricists, and both present non-factualist and anti-realist elements, they are motivated by different views in the philosophy of logic and mathematics.

Logic, Philosophy of Mathematics, and the Justification of Logic

S. Wagner noted that "[t]rying to characterise logic is pointless unless one has a viewpoint from which it matters what logic is". Both Hartry Field and Michael Resnik have a viewpoint from which it does. Field is very well known, among other things, for contending that mathematics is dispensable for science, for rejecting any compromise with mathematical entities. In order to show the viability of his view, he started the ambitious programme of reconstructing scientific theories substituting logic (second order logic) for mathematics, in particular, he develops a surrogate for the mathematics in Newtonian physics that has objective truth values (Field, 1980), one in which logic plays a fundamental role. One important feature this reconstruction is to have is that the sentences that obtain in his reconstruction in logical terms of scientific theories must have truth-values, some must be true, and they must be true independently of us. This is important because otherwise physical knowledge would not be objective knowledge, which Field claims it is. One problem for his view—one we will not be dealing with here—is technical: if second order logic is set theory, if mathematics is needed for logic, then his project fails. Similarly, he cannot accept that proof theory or model theory are adequate theories about logical facts because they use (or they are) mathematics. Thus, "being logical true" and "being valid" are for him primitive undefinable notions, notions about which speakers have a pre-theoretical grasp. As Resnik puts it: "it cannot be a necessary condition for something to be logically necessary that there exists a proof of it in some abstract formal language; similarly, it cannot be a necessary condition for a statement to be logically possible that it have a set theoretic model" (Resnik, 2000, pp. 182-3). Field takes both model-theory and proof-theory to be instruments that help us

*This work was supported by the Spanish Ministry of Science and Technology under project HUM2006-04955/FISO.

unveil our primitive notions of logical truth and logical consequence. Consequently, his reflections about logic are about what he calls 'our all purpose logic', not about logic understood as a mathematized discipline.

Resnik is a realist about mathematics, and a defender of logical anti-realism. He rejects Field's programme to reconstruct science without mathematics. He distinguishes 'logic' our correct inferential practice, from 'LOGIC', "our disciplined discussion of lower case logic", a mathematized discipline (Resnik, 2004). He is a realist about LOGIC because he is a realist about mathematics, while he is an anti-realist about logic. In particular, he rejects logical realism formulated in the following terms (Resnik, 2000, pp. 181):

1) claims about logical truth are true or false;
2) The truth values of such claims are independent of us, our linguistic conventions and inferential practices.
3) There are some sentences that are logically true.

Field explicitly distinguishes three levels in relation to logical notions (Field, 1996, pp. 370): i) the ground level of logical truth, ii) the level of logical validity, and iii) the level of logical necessity. Field is a factualist, a realist in relation to logical truth, while he is a non-factualist about validity, and about logical implication. On the other hand, Resnik distinguishes logical truth from logical validity as belonging to LOGIC, from their intuitive counterparts (logic), while he is a sceptic about logical necessity: logical truths are true, not *necessarily* true. As we said above, this conveys he is a realist about the mathematical notions of logical truth, logical consequence, logical implication, etc. On the other hand, he rejects realism about our all purpose logic: there is no fact of the matter about which logic is our all purpose logic because this is something one decides as a result of applying the methodology of wide reflective equilibrium.

When they discuss logical justification and the normative character of logic, they are both discussing logic understood as "our all purpose logic", "our inferential practice". In Field's case, this implies he is interested in the (paradox-generating) logic that underlies natural language. It also seems to imply that logic's main feature is going to be its *general* character, where the particular way in which logic is going to be general is by providing a *universal* canon for reasoning, one that is applicable not just to reasoning about this or that domain, but to all reasoning. In other words, he takes it that any appropriate theory about our all-purpose logic is to include a general truth predicate like the one we use in natural language. The reason is that the logic we take as correct sets limits on the way we move from a given set of beliefs to other beliefs; in particular:

> (I) If one knows [is certain] that A_1, through A_n together imply B then one's degree of belief in B should be such that $D(B) \geq D(A_1) + \ldots + D(A_n) - (n-1)$.

Now, no logic that includes a general truth predicate is generally truth-preserving. Thus, the implication relation defined by a logic that satisfies these features cannot be defined as being necessarily truth-preserving. As a result he claims that if we want to preserve the natural tie between validity and inferential practice, then

validity cannot be understood as necessary truth-preservation but as a normative notion (Field, 2009b, pp. 349-353).

Resnik agrees with Field that logic is to be normative in relation to our inferential practices (Resnik, 1985, pp. 223). He contends, against Thagard, for instance, that what logical theories are to model is not our inferential behaviour (with our empirical limitations), but the arguments we accept. From this point of view, logic is normative. The question is then what makes a given practice correct. Resnik's picture relies on our initial logical intuitions (not rational intuition), and in Goodman's methodology of wide reflective equilibrium applied to the case of logic. The logic (lowercase logic) we use, and the one we *should* use, is the one that allows us to attain wide reflective equilibrium:

Narrow reflective equilibrium seeks a fit between our intuitions concerning cases and our logical rules; wide reflective equilibrium permits us to go beyond such an initial fit and use either empirical or philosophical views to reform either our logical principles or our judgments about particular cases. (Resnik, 1985, pp. 225-6)

Internally one can modify one's views until attaining wide reflective equilibrium (Resnik, 1985, pp. 231), but there might be no reasons to argue for our choice in front of someone else's choice. Given different beliefs, endeavours, and intuitions different people might come to be in wide reflective equilibrium while defending that different logics, that is, different sets of logical truths, different notions of validity or of logical consequence are correct. Resnik considers we cannot "make sense of an ideal point where logicians are bound to agree at least concerning the logical data. […] for it is a matter of making them come to the same evaluations" (Resnik, 2000, pp. 189). He claims his view is non conventionalist and non relativist, a kind of "restrained logical non-cognitivism: sentences of ordinary language that seem categorically to attribute logical properties and relations are neither true nor false, and they actually perform other functions" (Resnik, 2000).

Both share the idea that there is no fact of the matter as to which logic one should use as our all purpose logic. But Resnik does not question that validity can be defined as truth preservation; yet, from his position it follows that someone, Field for instance, could come to be in wide reflective equilibrium by taking a compromise with the kind of logic he advocates and which conveys understanding implication in terms other than being necessarily truth-preserving.

Field's "Logicism" Assessed

Field takes logic is special, in contrast to other normative disciplines, in that most of us advocate as the goal of logic that of providing tools that guarantee necessary truth-preservation; the problem, as we have said above, is that there is no logic that can preserve truth in all generality: "this more general goal is one we can't consistently believe our logic to meet, unless we are to unduly restrict what counts as logic" (Field, 2009, pp. 356). But, he continues, "we *can* get a logic in which all the theorems are (necessarily) true, and I think we should want this since it's a special case of the goal of wanting the inferences to preserve truth when applied to

true premises. *And relative to this goal, many logics will genuinely disagree*: one logic will contain theorems whose truth will be rejected by proponents of the other logic." (Idem)

He also says that our all purpose logic:

> "… reduces to classical when certain assumptions are made. These assumptions seem plausible within ordinary mathematics and physics; so unlike Putnam's and Dummett's proposals, *there is no need to make modifications in ordinary mathematics and physics* (Field, 2008b, John Locke Lecture 3).

So, according to Field, in the case of mathematics and physics we can do without a logic that is general in the sense depicted above. Thus, the distinction he makes between the factuality of logical truths and the non-factuality of validity, allows him to combine his normative view about validity with his factualist view about which set of logical theorems is correct. Which logic is the logic that allows us to reconstruct physics is a factual issue because it has to do with the obtaining of a certain set of necessary truths.

The question is whether this view of his—namely that logical truth and logical validity are radically different—is *ad hoc* or not. Of course, the fact that no logic that includes a general truth predicate can include a characterization of itself as being truth preserving is not *ad hoc*; Results such as Curry's Paradox, Tarski's theorem or second Gödel's incompleteness theorem tell us that we cannot have it all. In such case we need to choose. But in order for his argument to succeed, he needs to convince us that only a logic that includes a general truth predicate of this sort is appropriate to set constraints on how we should move from one set beliefs to some other belief. His main argument in relation to this is that logic is to include a general truth predicate if it is to be our all purpose logic.

References

Field, H. (1980), *Science without Numbers*, Princeton, NJ, Princeton University Press.

— (2008), *Saving Truth from Paradox*, New York, Oxford University Press.

— (2008b), 'Logic, Normativity, and Rational Revisability', John Locke Lectures, Oxford, <http://www.philosophy.ox.ac.uk/lectures/john_locke_ lectures>

— (2009), 'Epistemology Without Metaphysics', *Philosophical Studies* 143, pp. 249-290.

— (2009b), 'Logical Pluralism', *The Review of Symbolic Logic* 2, n. 2, pp. 342-359.

— (2009c), 'What is the Normative Role of Logic', *Aristotelian Society Supplementary Volume* 83, n. 1, pp. 251-268.

Resnik, M. (1985), 'Logic: normative or descriptive? The ethics of belief or a branch of psychology', *Philosophy of Science* 52, pp. 221-238.

— (1996), 'Ought there to be but one logic?', in Copeland, B. J. (ed.), *Logic and Reality. Essays on the Legacy of Arthur Prior*, pp. 489-517.

— (2000), 'Against Logical Realism', *History and Philosophy of Logic* 20, pp. 181-194.
— (2004), 'Revising Logic', in Priest, G., Beall, J. C. y Armour-Garb, B. (eds), *The Law of Non-contradiction: New Philosophical Essays*, Oxford, Oxford University Press, pp.178-193.
Wagner, S. (1987), 'The Rationalist Conception of Logic', *Nôtre Dame Journal of Formal Logic* 28, pp. 3-35.

Significado y lógica. Sobre la función de la lógica según R. Brandom*

Nancy Núñez y Alessandro Moscarítolo
Universidad Central de Venezuela
nnunezm@gmail.com / moscaritolo.ucv@gmail.com

¿Es plausible concebir a la lógica como una herramienta vinculada con, dicho de la forma más general, nuestra manera de relacionarnos cognoscitivamente con el mundo? ¿Debe considerársela, más bien, una sucesión de manchas de tinta sin conexión alguna, en principio, con el modo en que tenemos experiencia del mundo y comunicamos esa experiencia? Este trabajo explorará una respuesta a esta pregunta, o a algo parecido a ella: la propuesta por Robert Brandom.

Brandom propone una teoría semántica edificada sobre una concepción de la racionalidad distinta a la ortodoxa racionalidad instrumental. El ejercicio de la "racionalidad expresiva" brandomiana consistiría en hacer explícitos los contenidos implícitos en nuestras creencias con el fin de someterlos a crítica, de insertarlos en el juego "de dar y pedir razones". Se trata de una concepción de la racionalidad que, afirma Brandom, sería afín al método socrático. Ahora bien, ¿en qué sentido se relaciona esto con una teoría del significado? En *Articulating reasons*, Brandom propone que esta última ha de desarrollarse como "una investigación del proceso de elucidación, del 'método socrático' de descubrir y reparar conceptos discordantes" (Brandom, 2000, p. 75). Sostiene, además, que la lógica desempeña un papel protagónico en la racionalidad reflexiva en general y, por lo tanto, en la teoría del significado en particular. Pero, ¿en qué está pensando Brandom al proponer una comprensión tal de la semántica filosófica? ¿Cuál es la relación entre ésta y la lógica?

En contraste con las concepciones semánticas que privilegian la referencia, hay quienes privilegian la inferencia como primitivo semántico. Los orígenes de esta posición pueden rastrearse hasta, por lo menos, los racionalistas modernos. Pero resulta de especial interés notar que representa la concepción semántica de Frege, del joven Frege de la *Conceptografía*. En efecto, el objetivo manifiesto de esta obra es aclarar contenidos conceptuales, y es claro que Frege propone explicar estos últimos en términos inferenciales: recuérdese cómo advierte que dos juicios, como "los griegos derrotaron a los persas en Platea" y "los persas fueron derrotados por los griegos en Platea", tienen el mismo contenido conceptual si "todas las inferencias que pueden extraerse del primer juicio, cuando a éste se lo combina con ciertos otros, también pueden extraerse del segundo cuando se lo combina con los mismos otros juicios" (Frege, 1879, sección 3, en Brandom,

* La realización de este trabajo fue posible gracias al financiamiento del CDCH-UCV. Agradecemos la valiosa colaboración de los profesores Ezra Heymann y Juan Rosales en su elaboración.

2000, p. 50) Por su parte, Brandom, quien también suscribe esta prioridad de la inferencia respecto de la referencia, la hereda directamente de Sellars. Entender un concepto es, para Sellars, "tener un dominio práctico respecto de las inferencias en las que está involucrado: saber, en el sentido práctico de poder distinguir (un tipo de *know how*), qué se sigue a partir de la aplicabilidad de un concepto, y a partir de qué se seguiría la misma" (Brandom, 2000, p. 48).

Ahora bien, podría asumirse espontáneamente que, cuando se habla de la constitución "inferencial" del contenido conceptual, se alude a inferencias formal o lógicamente correctas. Pero no es esto lo que piensa Brandom. Siguiendo a Sellars, propone dar primacía a las inferencias pre-lógicas o "materiales", a saber, aquéllas que son correctas únicamente en virtud de los *significados* de los conceptos involucrados. Por ejemplo, el contenido conceptual del enunciado "Valencia está al sur de Barcelona" implica *materialmente* "Barcelona está al norte de Valencia": este último puede inferirse del primero atendiendo únicamente al significado de "sur" y "norte". Entender el contenido conceptual de un enunciado presupone la capacidad de reconocer al menos algunas de las inferencias materialmente válidas que se siguen de él. Pero, y esto es central, no presupone la capacidad de reconocer inferencias *lógicamente* válidas. No sólo esto: Brandom insistirá en que es más bien la validez inferencial formal la que se explica en términos de la material, a la cual por tanto presupone (mientras que no es posible la explicación a la inversa).

Sin embargo, Brandom asigna al vocabulario lógico una función estelar no en la constitución, sino en la *explicitación* del contenido conceptual. Decíamos que Brandom defiende la importancia de un tipo de racionalidad distinta a la teórica y la práctica, la racionalidad expresiva, que consiste en hacer explícitas, bajo la forma de enunciados, las reglas seguidas en el pensamiento, el lenguaje y la acción, a fin de someterlas a crítica racional. Es en esta empresa donde el vocabulario lógico, y de forma emblemática el condicional, desempeñarían su función característica. Como el contenido conceptual está constituido inferencialmente, un enunciado del tipo $p \rightarrow q$ sería la forma de expresar las inferencias materiales asociadas a nuestro uso de un cierto concepto. Sellars expresa esta idea llamando "licencias inferenciales" a los enunciados condicionales. El disponer de tales licencias inferenciales hace posible, en efecto, la crítica racional de nuestro discurso y nuestras acciones: por ejemplo, permite someter a evaluación, y eventualmente rechazar, el uso de un concepto, cuando las consecuencias con las que nos comprometemos al usarlo nos resulten inaceptables.

Brandom sostiene que las otras conectivas de la lógica clásica, así como los operadores de la lógica modal, también pueden definirse en estos términos "expresivistas". Volveremos en un momento sobre esto, desde un enfoque ligeramente diferente.

Como Brandom, Frege en su *Conceptografía* también atribuye primacía al condicional. Al respecto, encontramos en Frege observaciones como ésta: "la relación hipotética definida con precisión entre contenidos de posibles juicios tiene una significación similar para la fundamentación de mi conceptografía a la

que tiene la identidad de extensiones para la lógica booleana" (Frege, 1979, p. 16, en Brandom, 2000, p. 59). Para Brandom, este énfasis en la inferencia, y no en la verdad o la referencia, es consecuencia del hecho de que el Frege de la *Conceptografía* en realidad no concibe a la lógica como una herramienta para probar la validez formal de argumentos, sino para aclarar, para hacer explícitos contenidos conceptuales. Brandom propone, entonces, que su concepción expresivista de la lógica tiene un antecedente directo nada menos que en Frege, y específicamente en una obra suya que es la partida de nacimiento de la lógica moderna.

En lo dicho hasta ahora sobre la subordinación de la inferencia lógicamente válida a la inferencia materialmente válida apenas se vislumbra un asunto que a Brandom, no obstante, le parece central. Hemos hablado sobre cómo el entender el contenido conceptual de un enunciado "presupone la capacidad práctica de reconocer al menos algunas de las inferencias materialmente válidas que se siguen de él", y de cómo el empleo posterior de la lógica sería un intento por codificar esa capacidad previa, por hacerla explícita. Es a este discurso sobre capacidades prácticas implícitas al que atenderemos ahora.

En una serie de conferencias dictadas a partir de 2006, Brandom ha ido exponiendo un programa filosófico que pretende mostrar "cómo el pragmatismo puede ser transformado de consejero pesimista e incluso nihilista de desesperanza teórica a un programa definitivo, sustantivo, progresivo y prometedor en la filosofía del lenguaje: en realidad, cómo puede entendérsele como simplemente la última fase del proyecto analítico" (Brandom, 2006, p. 1). Lo llama "pragmatismo analítico". Una de sus herramientas conceptuales es el llamado "análisis significado-uso", que responde a la idea de que toda relación semántica está pragmáticamente mediada. Entre otras, Brandom identifica las siguientes relaciones significado-uso: las llamadas "relaciones de suficiencia (y necesidad) PP", es decir, las relaciones entre dos conjuntos de capacidades prácticas que se establecen cuando un conjunto es condición suficiente (o necesaria) para el otro, y las "relaciones de suficiencia (y necesidad) PV", es decir las relaciones que se establecen cuando un conjunto de prácticas es condición suficiente (o necesaria) para el empleo de un cierto vocabulario.

¿Qué quiere decir que ciertas prácticas son condición suficiente o necesaria para otras prácticas? Brandom lo ilustra con la teoría sobre el funcionamiento de autómatas, porque "los autómatas son la encarnación práctica de algoritmos. Y los algoritmos generalmente dicen cómo puede ejercerse un conjunto de capacidades primitivas a fin de constituir capacidades más complejas" (Brandom, 2006, p. 2). En efecto, los llamados "autómatas de estados finitos" combinan sus capacidades primitivas de acuerdo con algoritmos condicionales, que especifican distintos tipos de respuesta frente al mismo estímulo, según el resultado de algo hecho previamente. En este sentido, elaboran capacidades complejas a partir de las simples y encarnan "relaciones de suficiencia PP". Pasando de los autómatas a los usuarios de conceptos, y de las capacidades prácticas en general a las capacidades prácticas específicamente *discursivas*, Brandom sostiene que hay dos capacidades cuyo manejo es condición al menos necesaria para todo aquello que merezca ser

considerado una práctica discursiva: la práctica de *aseverar* y la práctica de *inferir*. En otras palabras, la capacidad de participar en cualquier "juego de lenguaje" particular se explicaría como una capacidad "compleja", en el sentido de elaborada algorítmicamente a partir de las capacidades "primitivas" de aseverar y distinguir las relaciones inferenciales en que se encuentran dichas aseveraciones. En la jerga de Brandom, las últimas serían suficientes PP para la primera. Así, la capacidad de participar en aquel juego de lenguaje que involucra emplear locuciones condicionales constituiría una capacidad subsidiaria de capacidades cognoscitivas y lingüísticas más fundamentales. La explicación mediante elaboración algorítmica se resumiría como sigue. Para comenzar, todo sistema capaz de prácticas discursivas puede, por definición, aceptar o rechazar la inferencia de *q* a partir de *p*. También, por definición, puede aseverar *p* por una parte, y *q* por la otra. Así las cosas, puede enseñársele al sistema a aseverar "si p entonces q" en aquellos casos en los que esté dispuesto a responder en la práctica a la inferencia de *q* a partir de *p* considerándola correcta.

En esta segunda línea de argumentación se hace evidente otra consecuencia ya familiar: el papel expresivo de las expresiones condicionales. Apreciamos aquí que el condicional hace explícito, en formato lingüístico, lo que en principio son sólo capacidades *prácticas* de distinguir inferencias materiales correctas. Recurriendo a su "análisis significado-uso", Brandom expresa ahora este papel expresivo señalando que el vocabulario constituido por las locuciones condicionales se halla en la relación *LX* con respecto a la práctica de distinguir inferencias correctas: es *elaborado* a partir de ella (L) y *explicativo* de la misma (X).

Como era de esperarse, Brandom propone que, además del condicional, en general el vocabulario lógico es un vocabulario LX, es decir, un vocabulario caracterizado por desplegarse con base en un conjunto de prácticas "complejas", en el sentido de elaboradas algorítmicamente a partir de prácticas "primitivas", necesarias para toda práctica discursiva, y con el poder de especificar explícitamente estas últimas.

Veamos cómo se definiría la negación lógica de este modo. Brandom argumenta que, además de las capacidades prácticas de aseverar e inferir, otra práctica necesaria para cualquier práctica discursiva consiste en la capacidad de distinguir afirmaciones materialmente incompatibles. A partir de esta capacidad se puede explicar, mediante elaboración algorítmica, la posesión de la capacidad "compleja" requerida para el despliegue de la pieza de vocabulario que es la negación lógica. Y, una vez desplegada, la negación lógica, junto con el condicional, permiten hacer explícita esa práctica previa de distinguir afirmaciones incompatibles: "Si la mesa es rectangular, entonces *no* es circular".

En realidad, Brandom termina asignándole gran importancia a la noción de incompatibilidad en su explicación del papel expresivo de la lógica, al punto de proponer una "semántica de la incompatibilidad", que permitiría definir todas las conectivas de la lógica clásica, más los operadores modales de necesidad y posibilidad, en términos de relaciones de incompatibilidad. Por razones de espacio

nos limitamos sólo a mencionar esta propuesta. De todos modos, está articulada en el mismo espíritu de la concepción de la lógica que hemos explorado aquí. Frente a la pregunta del comienzo, es obvio ahora cuál será la respuesta de Brandom; una respuesta que, como él mismo advierte, "mantiene la esperanza de recuperar para el estudio de la *lógica* una significación directa para proyectos que han estado en el corazón de la *filosofía* desde su comienzo socrático" (Brandom, 2000, p. 77).

Referencias bibliográficas

Brandom, R. (2000), *Articulating reasons. An introduction to inferentialism*, Cambridge, Harvard University Press.

— (2006), 'Elaborating abilities. The expressive role of logic', disponible en línea: <http://www.pitt.edu/~brandom/locke/locke-w2.html>

Frege, G. (1879), *Conceptografía*. México, UNAM, 1972

— (1979), *Posthumous Writings*, Chicago, University of Chicago Press.

Un argumento a favor de la existencia de (verdaderas) lógicas paraconsistentes

Carlos A. Oller
Universidad de Buenos Aires / Universidad Nacional de La Plata
coller@ciudad.com.ar

Introducción

En la literatura es posible encontrar una multiplicidad de lógicas paraconsistentes, i.e. de sistemas que no validan la regla del *ex contradictione quodlibet* (*ECQ*). Sin embargo, se ha argumentado que no es posible que exista una verdadera lógica paraconsistente porque una negación para la que no valga el *ECQ* no puede ser una verdadera negación. Se arguye que una verdadera negación es un operador formador de contradictorios y que las negaciones de las lógicas paraconsistentes no pueden serlo.

En este trabajo expondremos el argumento en contra de la posibilidad de la existencia de verdaderas lógicas paraconsistentes y la respuesta de F. Paoli (2003) a este argumento. Además, presentaremos un nuevo argumento —inspirado, como el de Paoli, en una sugerencia de Susan Haack— en contra de la tesis de la imposibilidad de la existencia de verdaderas lógicas paraconsistentes.

El argumento en contra de la existencia de (verdaderas) lógicas paraconsistentes

La caracterización formal más comúnmente aceptada de las lógicas paraconsistentes considera que la paraconsistencia es una propiedad de la relación de consecuencia de esas lógicas. Sea $\models$ una relación de consecuencia, caracterizada en términos sintácticos o semánticos. La relación de consecuencia $\models$ es explosiva si y sólo si, para toda fórmula A y B, se da que $\{A, \neg A\} \models B$, un patrón inferencial al que se suele denominar *ex contradictione quodlibet* (*ECQ*). Una relación de consecuencia es paraconsistente si y sólo si no es explosiva. Una lógica es paraconsistente si y sólo si su relación de consecuencia es paraconsistente, *i.e.* si no valida la regla del *ex contradictione quodlibet*. Esta caracterización de las lógicas paraconsistentes está, pues, basada en un criterio puramente negativo.

Una posición filosófica hostil a la lógica paraconsistente, que haría de la paraconsistencia una propiedad vacía, es la que tiene su fuente en el *dictum* de Quine "cambio de lógica es cambio de tema" (Quine, 1970). La idea fundamental de esta crítica a la paraconsistencia es que no es posible tener una lógica con una negación que sea una verdadera negación y que, al mismo tiempo, no satisfaga el principio del *ECQ*. Como en lógica elemental un cambio en la teoría —como el

que no valga el *ECQ* para la negación— es un cambio de significado, el lógico paraconsistente sólo cambia de tema al cambiar la teoría lógica clásica.

Una forma restringida de esta crítica fue dirigida por Priest y Routley (1989) contra la lógica paraconsistente de da Costa. La negación de da Costa (1974) no sería, según estos autores, una verdadera negación porque la (verdadera) negación es un operador formador de contradictorios y la negación paraconsistente de da Costa es sólo un operador formador de subcontrarios. Recordemos que dos fórmulas A y B son subcontrarias si $(A \vee B)$ es una verdad lógica, y contradictorias si $(A \vee B)$ es una verdad lógica y $(A \wedge B)$ es lógicamente falsa. Si $(A \wedge \neg A)$ no es lógicamente falsa, como sucede en la lógica de da Costa, entonces la negación simbolizada con $\neg$ no es un operador formador de contradictorios y, por lo tanto, tampoco es una verdadera negación. En un corto y muy citado artículo, Slater (1995) extiende esta crítica al sistema trivalente LP de Priest (1979) porque, aunque $(A \wedge \neg A)$ no recibe nunca el valor de verdad designado "(sólo) verdadero" en este sistema, tanto A como $\neg A$ pueden recibir el otro valor designado "tanto verdadero como falso". Pero, no puede suceder que dos fórmulas contradictorias reciban el valor verdadero, sea éste el valor "(sólo) verdadero" o el valor "tanto verdadero como falso". La conclusión de Slater es que A y $\neg A$ pueden no ser contradictorios en LP y que, por lo tanto, la negación de este sistema no es una verdadera negación.

Argumentos a favor de la existencia de (verdaderas) lógicas paraconsistentes

Francesco Paoli ofrece en su artículo "Quine and Slater on Paraconsistency and Deviance" (Paoli, 2003) una respuesta al argumento de Slater desarrollando una idea sugerida por Susan Haack en su *Deviant Logic* (1974). En efecto, Haack trae a cuento la formulación de Gentzen de la lógica minimal, que difiere de la lógica clásica sólo en lo que respecta a las reglas estructurales pero coincide con ella en lo que hace a las reglas para las conectivas. Como las reglas estructurales no contienen ninguna referencia a las conectivas, la divergencia entre estos sistemas parece poder explicarse sin apelar a un cambio en el significado de las conectivas.

Paoli reformula y desarrolla el argumento de Haack en el marco de una semántica para la lógica basada en la teoría de la demostración (*proof-theoretic semantics*) (Wansing, 2000). De acuerdo con ella, el significado de una constante lógica c queda especificado por las reglas de secuentes que nos permiten introducir la conectiva —las reglas operativas que fijan lo que Paoli llama *el significado operativo de c*— y por las reglas estructurales —que fijan lo que llama *el significado global de c en el sistema*. Los principios e inferencias que involucran a una conectiva c y que resultan sancionados por el sistema dependen no sólo del significado operativo de c sino también de su significado global. Si no hay un cambio en el significado operativo de una constante c al pasar de un sistema S_1 a otro S_2 que valida distintos principios lógicos para c en virtud de tener diferentes reglas estructurales, entonces hay una rivalidad genuina entre esos sistemas. En el caso de la negación muchas lógicas divergentes tienen, en su presentación como cálculos de secuentes, las mismas reglas operativas que la lógica clásica para esta

conectiva y, por lo tanto, la negación tiene el mismo significado operativo en todos estos sistemas y es, por así decirlo, una verdadera negación en todos ellos.

En lo que sigue presentaremos otro argumento a favor de la existencia de verdaderas lógicas paraconsistentes en términos de la semántica estándar para la lógica elemental que, como el argumento de Paoli, desarrolla la sugerencia de Haack mencionada más arriba. La idea intuitiva es que si las conectivas de un sistema de lógica proposicional quedan caracterizadas por las cláusulas de la semántica veritativo-funcional de la lógica proposicional clásica, pero la noción de consecuencia lógica se caracteriza de manera no estándar, entonces estamos frente a una verdadera divergencia que no consiste en un mero cambio de significado de las constantes lógicas. En efecto, el hecho de que la afirmación metalingüística $\Gamma \models A$ sea verdadera, para algún Γ y A, en un sistema y falsa en otro puede deberse no sólo a un cambio en el significado de las constantes lógicas que aparecen en las fórmulas de Γ y en A, sino también a un cambio en el significado $\models$. Si es posible explicar las diferencias en el conjunto de inferencias válidas entre dos sistemas apelando a una diferencia en su noción de consecuencia lógica, al tiempo que la caracterización semántica de las constantes lógicas permanece igual, entonces estaremos en presencia de una verdadera divergencia y no de un simple cambio de significado de las conectivas.

Un ejemplo de un sistema proposicional de este tipo es la lógica de la implicación analítica de Parry (1933, 1989) cuando la implicación analítica no se trata —como lo hace Parry— como una conectiva del lenguaje sino como una relación de consecuencia lógica (Oller, 1999). Parry quiere caracterizar la relación de "implicación real", es decir la relación entre dos proposiciones que es necesaria y suficiente para permitirnos pasar, en virtud de razones puramente lógicas, de la afirmación de la primera proposición a la afirmación de la segunda. Es necesario apuntar al pasar que, aunque la intención de Parry era ofrecer una teoría de la consecuencia lógica, concibe su noción de implicación, siguiendo el ejemplo de C. I. Lewis, como una conectiva del lenguaje en lugar de tratarla como una relación metalingüística entre expresiones del lenguaje. La caracterización de esa noción debería, en la opinión de Parry, evitar paradojas de la implicación estricta como que *q* implique *p o no-p*, o que *p y no-p* implique *q*; Parry añade al rechazo de esas implicaciones la objeción al principio de adición, es decir a que *p* implique *p o q*. En todos los casos debe evitarse que en el consecuente de una implicación aparezcan términos irrelevantes respecto de su antecedente.

M. Dunn (1972) propuso en una semántica algebraica para la implicación de Parry en la cual un modelo para un sistema proposicional de implicación analítica es un par ordenado $<v, s>$, donde v es una valuación booleana y s es una función que asigna a cada fórmula bien formada un subconjunto de un conjunto no-vacío S, que puede considerarse como un conjunto de unidades de contenido. La función s de asignación de contenidos cumple con la siguiente cláusula S: $s(A) = \cup\{s(p)$: p es atómica y aparece en $A\}$; es decir, el contenido de una formula es la unión del contenido de las variables proposicionales que aparecen en ella. El contenido de un conjunto de oraciones Γ es la unión de los contenidos de sus miembros.

Diremos que una proposición A es una consecuencia analítica de un conjunto de premisas Γ si y sólo si A es verdadera bajo toda valuación booleana v que haga verdadera a todos los miembros de Γ, y, dada cualquier asignación de contenidos s que cumpla la cláusula S, el contenido de A resulta incluido en el contenido de Γ. Se desprende de esta caracterización semántica de la relación de consecuencia analítica no satisface la regla del *ex contradictione quodlibet*. La lógica resultante es, pues, una lógica paraconsistente en la que las conectivas quedan caracterizadas por las cláusulas semánticas estándar y en la que, por lo tanto, la divergencia respecto de la lógica clásica no se debe a un cambio de significado de las constantes lógicas sino a un cambio en la noción de consecuencia lógica. En particular, la negación de esta lógica tiene las mismas condiciones de verdad que la negación clásica, y dos fórmulas contradictorias A y $\neg A$ no pueden ser ambas verdaderas de acuerdo a la semántica de este sistema. La regla del *ECQ*, pues, no resulta inválida debido a un cambio de significado de la negación, sino a una concepción de la consecuencia lógica que no es la clásica.

Referencias bibliográficas

da Costa, N. C. A. (1974), 'On the theory of inconsistent formal systems', *Notre Dame Journal of Formal Logic* 15, pp. 497-510.

Dunn, J. M. (1972), 'A modification of Parry's analytic implication', *Notre Dame Journal of Symbolic Logic* 13, pp. 195-205.

Haack, S. (1974), *Deviant Logic*, Cambridge, Cambridge University Press.

Oller, C. A. (1999), 'Paraconsistency and Analyticity', *Logic and Logical Philosophy* 7, pp. 91-99.

Parry, W. T. (1933), 'Ein Axiomensystem für eine neue Art von Implikation (analytische Implikation)', *Ergebnisse eines mathematischen Kolloquiums* 4, pp. 5-6.

Parry, W. T. (1989), 'Analytic Implication: Its History, Justification and Varieties', en J. Norman y R. Sylvan (eds.), *Directions in Relevant Logic*, Boston, Kluwer, pp. 101-118.

Priest, G. G. y Routley, R. (1989), *Paraconsistent Logics, Essays on the Inconsistent,* Munich, Philosophia Verlag.

Paoli, F. (2003), 'Quine and Slater on Paraconsistency and Deviance', *Journal of Philosophical Logic* 32, pp. 531–548.

Priest, G. (1979), 'The logic of paradox', *Journal of Philosophical Logic* 8, pp. 219-241.

Quine, W. V. O. (1970), *Philosophy of Logic*, Englewood Cliffs, Prentice-Hall.

Slater, B. H. (1995), 'Paraconsistent logics?', *Journal of Philosophical Logic* 24, pp. 451-154.

Wansing, H. (2000), 'The idea of a proof-theoretic semantics', *Studia Logica* 64, pp. 3-20.

Boxes and explosions

Andreas Pietz
Universitat de Barcelona
andreas.pietz@gmail.com

Introduction

In the discussion with David Lewis and others about how we reason in the face of inconsistency, Graham Priest reverted to a very interesting strategy (1999): He wrote a short story, "Sylvan's Box". That story features an inconsistent object, a box that is both empty and not empty. Priest then argued that the logic that the reader has to adopt to grasp what is going on in the story cannot be classical. Such reasoning follows the rule of *Ex Contradictione Quodlibet* aka *Explosion*: Anything follows from a contradiction. Classical logic would lead the reader to infer things that are clearly not true in the story.

Any logic that does not allow inferences along the lines of Explosion is called a paraconsistent logic[1]. Priest's story shows that we can reason about inconsistent situations, and that we use a paraconsistent logic to do so. I will argue that this strategy has been very successful, and that in fact it is even more successful than Priest has claimed. Let me start out by giving a quick summary of the story.

Sylvan's Box

Priest's largely autobiographical story goes as follows: After the sudden death of his long time friend and colleague, Richard Sylvan (Routley), Priest drives to Sylvan's remote home to start working through his *Nachlass*. In between piles of papers he finds a little box with the words "Inconsistent Object" written on it. As he inspects it he finds it, incredibly enough, both empty and not empty. Inside there is a little figurine and, at the same time, there is no figurine whatsoever. This is not an illusion, the box is an object with perfectly contradictory properties. Priest and Nick Griffn, who is also present, discuss what to do with such a remarkable find, and they find themselves in two minds. The story ends on an extremely inconsistent note: Priest takes the box with him when he leaves Sylvan's house, while Griffin buries it in Sylvan's garden.

After this truly inconsistent ending, the reader is presented with a questionnaire about the story that tests whether the reader has understood what is going on in the story. If he answers "Yes" to the question whether the box was shot off to the

1. In other words, there is not one logic that is *the* paraconsistent logic, as is sometimes mistakenly assumed, but rather a whole family of them. Another important source of confusion is that paraconsistency is often erroneously taken to be the thesis that there are true contradictions. This metaphysical doctrine is called dialetheism. Dialetheism presupposes paraconsistency, but not vice versa.

moon, we will say that he has not understood the story at all. But that shows that the correct logic to utilize to reason about the story cannot be classical, because Explosion would license the inference from "The box is empty" and "The box is not empty" to "The box is shot off to the moon."

However, neither would the reader have understood what was going on by following Lewis's suggestion how to deal with inconsistent bodies of information in (1982). Here the idea would be to break up the story into (maybe maximally) consistent chunks and only draw inferences from those. But that way the reader would not only miss the whole point of the story, he would be drawing false inferences. Asked why Priest was so excited in the story, he would have to either answer it was because he found the box to be empty, or alternatively answer that it was because the box wasn't empty. But neither is right, Priest was excited because the box was both empty and not empty.

Seeing that even David Lewis came to appreciate this point, one might try to make all this into a very strong argument for paraconsistent logic. At the very least, we have to be able to employ a paraconsistent logic to deal with this story, and we don't seem to have much of a problem with that. If the logic we are employing is not paraconsistent from the outset, then it seems strange that we can make the transition so easily.

Logics and Stories

Priest actually doesn't take that line of argument. He suggests that for any logic whatsoever, a story could be told that would make it seem the only suitable logic for this story. That is what he did with Sylvan's Box for paraconsistent logic, and he suggests one could write a story about a man having entered a room through two doors without having entered through one in particular, a story that is supposed to get the reader to reason along the lines of quantum logic. It is interesting to follow this hypothesis through and investigate what kinds of logic can be forced onto the reader in a similar way. I will write that a story *induces* a given logic when this kind of manouver succeeds.

How would a story look that induces fuzzy logic, or one that induces intuitionistic logic? I think these are fascinating questions, but in this paper I'd like to experiment with a seemingly easier case, namely classical logic.

Inducing Classical Logic

So, how might we go about telling a story that induces classical logic? In other words, how would a story look that would force someone who believes that we normally reason paraconsistently to draw inferences along the lines of *Explosion*?

In order to induce *Explosion*, the story has to sport an inconsistency, and a quite blatant one at that. That's because we cannot make out a difference in the reasoning of someone who is employing classical logic and someone who is reasoning paraconsistently if there is not contradiction in sight.

We already have a good inconsistent story to work with, namely *Sylvan's Box*. Let's try to turn it into a story that induces *Explosion*. As *Explosion* tells us that

anything whatsoever follows from a contradiction, one might simply try to append all kinds of random statements to the story. E.g.:

> Alternative Ending 1:
>
> *"[Sylvan's Box]*
>
> *As I [ie Priest, the story is told from the first person perspective] drove down into town with the box in the trunk of my car, I came about a gaggle of pigs flying around my car, and the moon was made of green cheese. To my further surprise I realized that the sum of five and seven was fourteen. And eleven as well."*

Append enough of these, and a logician might get what is meant to be going on, but surely this isn't much more than an insider joke. No layman reading the story would take those statements to be conclusions of the former contradiction. The reason is simply that one cannot make statements be taken as consequences of each other simply by writing them in sequence.

Then what does it take to turn a series of statements into a logical sequence from premises to conclusions? I have to say that I can't think of any other way to pull this off than to state the aim outrightly, for example thus:

> Alternative Ending 2:
>
> *"[Sylvan's Box]*
>
> *As I drove down into town, I checked the contents of the box again. Sure enough, it was still empty. And sure enough the box still contained the little figurine. Therefore, a gaggle of pigs turned up on the horizon and flew towards me, soon to be circling around my car."*

I suggest that this strategy will not be successful, even if it was executed with more literary taste than these attempts. To my mind, the problems it runs into are the same that the literature knows under the name *imaginative resistance*. In short, the reader, even if she has swallowed all that preceded the last sentence, is not likely to buy this last part. It's not that pigs can fly that she can't accept, but rather that this is supposed to be a logical consequence of the preceding contradiction. It's exactly the "therefore" that will upset her reading experience.

Imaginative resistance

The term "imaginative resistance" came up in the discussion about the problems that arise in relating deviant morality or humor in stories. A story might have a mob of outraged commuters killing a couple that is clogging up traffic on the freeway (example from Weatherson, 2004). It's no problem at all to tell this in a way that engages the reader enough to have him "buying" it, i.e. to get him to hold the given events as true in the story. However, it is very hard indeed to do the same with a moral judgment about the situation. Just because the narrator tells you that the mob was right to kill the couple doesn't make it true, not even true in the story. The reader's imagination will revolt at that point, hence the name *imaginative resistance*. In Weatherson's words, "we refuse, fairly systematically, to play along with the author here"(2004, p. 2).

Resisting classical logic

What I'd like to suggest is that the failure of the examples I gave above to convince the reader is of the same kind as the cases where deviant moral judgements or divergent senses of humor are attempted to be pushed onto the reader. It would be interesting to see which of the explanations that have been offered in the moral cases might carry over to the logical case, but my point is sufficiently made by flagging the phenomenon. I think that the reader of one of my proposed stories will respond in exactly the same way as the reader of Weatherson's story about the avenging commuter. He might accept all that happens factually, but when it comes to judging what follows from what, he will insist on doing the inferring for himself and reject any inferential moves the narrator makes that he wouldn't have made himself.

Now, a last obvious observation: It is of course only when the moral judgements of the narrator clash with our own that we resist them. If Weatherson's narrator had gone on to tell us that it was wrong of the man to kill two people just because they were causing a traffic jam, it might have struck us as a strange bit of storytelling. However, we wouldn't have resisted the judgement and concluded that the murder was indeed justified, just because we don't like to be told what to think. That would be a bit juvenile, after all.

If the case of logical resistance indeed is of the same sort as the other kinds, then it would be hard for anyone to argue that a logic that we feel imaginative resistance toward is actually the one that we employ outside of story-reading. It would be the only case in which something like this is happening.

So where, does that leave us? Priest has shown with "Sylvan's Box" that we are perfectly able to reason paraconsistently. I argued that a similar move is very hard to pull off for the classical logician, because we don't like to be told to reason classically, just like we don't like to be told that pointless murder is right. Insisting now that classical logic is nonetheless the logic underlying our thinking seems more than a bit dogmatic.

References

Lewis, D., (1982), 'Logic for Equivocators', *Noûs* 16, n. 3, pp. 431-441.

— (1978), 'Truth in Fiction', *American Philosophical Quarterly* 15, pp. 37-46.

Priest, G., Beall, J. C. and Armour-Garb, B. (eds.) (2004), *The Law of Non-Contradiction: New Philosophical Essays*, Oxford, Oxford University Press.

Priest, G. (1999), 'Sylvan's Box: A Short Story and Ten Morals', *Notre Dame Journal of Formal Logic* 38, pp. 573-582.

— (2006), *Towards Non-Being: the Logic and Metaphysics of Intentionality*, Oxford, Oxford University Press.

Weatherson, B. (2004), 'Morality, Fiction, and Possibility', *Philosopher's Imprint* 4, n. 3, pp. 1-27.

From Hilbert's mathematical problems to Gödel's program and beyond

Joan Roselló
Universitat de Barcelona
joanrosello@gmail.com

From Hilbert to Gödel

In the year 1900, David Hilbert gave an address at the meeting of the Second International Congress of Mathematicians. The title of Hilbert's lecture was simply "Mathematische Probleme" and consisted of a list of 23 unsolved –by this time– problems, which led to a considerable amount of important research in different fields of mathematics. Three of them involved mathematical logic and the foundations of mathematics and their solutions were to have a deep impact on their future development. These are the problems numbered 1, 2 and 10. Hilbert's First Problem called for a proof of Cantor's *Continuum Hypothesis* (CH). Hilbert's Second Problem asked for a direct proof of the consistency of the axioms which determine the field of the real numbers. Finally, Hilbert's Tenth Problem prompted to devise an algorithm for the determination of the solvability of any diophantine equation. The great logician and mathematician Kurt Gödel (1906-1978) contributed in a decisive way to the solution of the three problems.

In respect to the first problem, Gödel (1938) established in that if ZFC (Zermelo-Fraenkel set theory with the axiom of Choice) is consistent, then so is ZFC+CH. The consistency of ZFC+$\neg$CH was established by Paul Cohen in 1963 by a method called *forcing*. So, by Gödel's and Cohen results, CH was finally shown to be independent of the axioms of ZFC.

With regards to Hilbert's second problem, Gödel (1931) demonstrated in the incompleteness of every consistent and sufficiently strong theory T (such as PA or ZFC) and soon after he also showed informally the unprovability in such a theory of the statement formalizing "T is consistent". This yielded a negative solution to Hilbert's second problem for it established the impossibility for even weak theories like *Peano arithmetic* (PA) of demonstrating his own consistency.

Regarding the third problem, Gödel also gave the first step towards Matiyasevic's (negative) solution of Hilbert's tenth problem. For Gödel did not only introduce the general notion of recursiveness, but he also demonstrated that every recursive function can be defined by a finite number of existential and bounded universal quantifiers (which is essential in the Matiyasevic-Robinson-Davis-Putnam proof of the unsolvability of Hilbert's tenth problem).

Undecidable sentences and Gödel's program

Shortly after Gödel had published his famous result on the incompleteness of arithmetic, he wrote this about his theorem:

> The procedure just sketched furnishes, for every system that satisfies the aforementioned assumptions, an arithmetical sentence that is undecidable in that system. That sentence 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). (Gödel, *1931?, p. 35).

We know, for example, that moving from arithmetic to analysis (second order arithmetic) or set theory we can prove $Cons_{PA}$ (the number theoretic translation of the statement that PA is consistent). 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. The first is obviously CH, but there are other sentences from set theory that are independent of ZFC and could be included in the class of absolutely undecidable sentences such as, for example, Suslin's hypothesis or the hypothesis that every set is constructible ($V = L$ for short).

It is well known that Gödel proved that if ZFC is consistent, then $\mathrm{ZFC} + V = L$ is also consistent and the later implies CH. Gödel himself considered in his *1939b (p. 155) and other writings of the same period that the statement $V = L$ was absolutely undecidable, but in his 1946 (p. 151) he came to believe that there might be no absolutely undecidable sentences and, therefore, that $V = L$ was surely not absolutely undecidable.

At around the same time Gödel also conjectured in an expository article about CH (1947) that this statement was not only consistent with ZFC, as he had previously shown, but actually independent, a conjecture that was confirmed by 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" (Gödel, 1947, p. 181).

Therefore, according to Gödel, in order to settle these undecidable sentences in ZFC such as CH or $V = L$, we should extend ZFC with new axioms and more precisely with the so called *axioms of infinity* or *large cardinal axioms*, *i.e.*, "propositions asserting the existence of very great cardinal numbers or (which is the same) of sets having these cardinal numbers" (*Ibid.*, p. 182).

But because all large cardinal axioms known by those times (inaccessible and Mahlo's cardinals) failed to settle CH, since they were all consistent with $V = L$, Gödel proposed the search of "other (hitherto unknown) axioms of set theory"

whose acceptance 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.*, pp. 182-183).

Does mathematics need new axioms?

A lot of work has since been done in the search of new axioms for set theory that can solve the continuum problem. Nonetheless, the most known of these axioms are claims about the existence of large cardinals (measurable, Woodin, supercompact, etc.) or are implied by them (the *Projective Determinacy* axiom, for example, is a consequence of the existence of infinitely many Woodin cardinals). In this way, *Gödel's program* for extending ZFC with new axioms in order to settle hitherto undecidable sentences (in ZFC) such as CH has been identified with the so called *large cardinals program.*

However, even the program for large cardinals has been very successful "below" CH (see Koellner, 2006, p. 172) for a precise statement of this fact), it has not been capable of proving or disproving CH itself. This breaking down of the large cardinals program and any other approaches to settle CH has brought some authors to cast doubts about the possibility of deciding CH and the need of new axioms for mathematics.

S. Feferman has argued, for example, "that the Continuum Hypothesis is […] an "inherently vague" statement, and that the continuum itself, or equivalently, the power set of the natural numbers, is not a definite object" (Feferman, 2000, p. 405). It follows from this "that the conception of the whole of the cumulative hierarchy (…) is even more so inherently vague, and that one cannot in general speak of what is a fact of the matter under that conception" (*Ibid.*). This includes obviously large cardinal axioms.

Even though Gödel's program for searching new axioms for set theory in order to settle CH has not been entirely successful, what about his prediction that higher level axioms will help us to decide number-theoretic statements of genuine mathematical interest such as, for example, Riemann's Hypothesis? (Gödel, *1951, p. 307). This is an important question because the solution of a previous formulated number-theoretic problem with the aid of large cardinal axioms would be surely a decisive factor in favour of their acceptance by the mathematical community.

We know indeed, as remarked by Gödel itself, that "each of these set-theoretical axioms entails the solution of certain diophantine problems which had been undecidable from the previous axioms" (*Ibid.*). But the truth is that the kind of diophantine problems which have been showed to be solvable are quite far from the problems which have preoccupied mathematicians in the last two or three centuries. Moreover any open arithmetical problem (such as Riemann Hypothesis or Goldbach conjecture) or finite combinatorial problem of genuine mathematical interest has been solved with the aid of such axioms.

A common argument against this claim is the fact that H. Friedman has been capable of producing some finite combinatorial statements φ whose proof require the existence of large cardinal axioms (Mahlo type and even stronger). But as Feferman has pointed out, if we call *S* to the system obtained by adding to ZFC the axiom of infinity needed for the proof of φ, then the *truth* of φ depends essentially on accepting 1 – Con (*S*) and hence is *not* a result of ordinary mathematical reasoning. So "it is begging the question to claim that this shows we need axioms of large cardinals (…) since this only shows that we "need" their 1-consistency" (Feferman, 2000, p. 407).

Concluding remarks

Can we conclude after all that mathematics doesn't need new axioms? Even though there is no intrinsic justification of large cardinal axioms beyond Mahlo's (see Koellner, 2006, pp. 162-166, for a discussion) and no evidence for their practical need in order to settle open arithmetical and finite combinatorial problems, we can still look for an extrinsic justification of their need in other areas of mathematics.

J. Steel has argued, for example, that large cardinal axioms "have proved crucial to organizing and understanding the family of possible extensions of ZFC" (Steel, 2000, p. 425) and that they have solved "all the questions about projective sets from classical descriptive set theory" (*Ibid.*, p. 428). Regarding this, K. Hauser has also pointed out that although "the first two levels of the projective hierarchy (for which a structure theory is already obtainable in ZFC) suffice for the needs of "ordinary" mathematical practice", we can find "examples of higher-level projective sets arising naturally in analysis and topology" (Hauser, 2002, p. 275).

In another direction, H. Friedman has predicted that because of the interaction of his new Boolean Relation Theory (BRT) and related developments, which provide compelling uses for Mahlo's and larger cardinals, with nearly all areas of mathematics, "large cardinal axioms will begin to be accepted as new axioms for mathematics" (Friedman, 2000, p. 438). The success of large cardinal axioms will depend therefore on their use in some area of mathematics with *essential* interconnections with other core areas of mathematics (which is not the case of set theory (*Ibid.*, p. 435). Only the future will tell us whether BRT will play this central role in normal mathematics predicted by Friedman.

References

Feferman, S. (2000), 'Why the programs for new axioms need to be questioned', *Bulletin of Symbolic Logic* 6, pp. 401–413.

Friedman, H. (2000), 'Normal mathematics will need new axioms', *Bulletin of Symbolic Logic* 6, pp. 434–446.

Gödel, K. (1931), 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38, pp. 173–198, in [Gödel (1986), pp. 144–195].

— (*1931?), 'On undecidable sentences', in [Gödel (1995), pp. 31-35].
— (1938), 'The consistency of the axiom of choice and of the generalized continuum hypothesis', *Proc. Natl. Acad. Sci. U.S.A.* 24, 556-557, in [Gödel (1990), pp. 26-27].
— (*1939b), 'Lecture at Göttingen', in [Gödel (1995), pp. 127-155].
— (1946), 'Remarks before the Princeton bicentennial conference on problems in mathematics', in [Gödel (1990), pp 150-153].
— [1947], 'What is Cantor's continuum problem?', *The American Mathematical Monthly* 54, pp. 515-527, in [Gödel, (1990), pp. 176-187].
— (*1951), 'Some basic theorems on the foundations of mathematics and their implications', in [Gödel (1995), pp. 304-323].
— (1986), *Collected Works*, *Vol. I. Publications 1929–1936*, edited by S. Feferman *et al.*, New York, Oxford University Press.
— (1990), *Collected Works*, *Vol. II. Publications 1938-1974*, edited by S. Feferman *et al.*, New York, Oxford University Press.
— (1995), *Collected Works*, *Vol. III. Unpublished Essays and Lectures*, edited by S. Feferman *et al.*, New York, Oxford University Press.
Hauser, K. (2002), 'Is Cantor's Continuum Problem Inherently Vague?', *Philosophia Mathematica* 10, pp. 257-285.
Koellner, P. (2006), 'On the Question of Absolute Undecidability', *Philosophia Mathematica* 14, n. 2 (2006), pp. 153–188.
Steel, J. R. (2000), 'Mathematics needs new axioms', *Bulletin of Symbolic Logic* 6, pp. 422–433.

The practice of judging information-theoretic validity and invalidity: Heuristics, examples, and queries*

José M. Sagüillo
University of Santiago de Compostela
josemiguel.saguillo@usc.es

An epistemic gap in the current logical paradigm

I want to address the issue of how judgment of logical validity and logical invalidity arises in the real field of mathematical practice. The rationale of methodological practice imposes the prior experiences of establishing logical implication and logical independence before these can be adequately captured and codified in a logical theory. Thus, "practice precedes theory" and logic is no exception but rather a golden example to this motto: the practice of proof and disproof is previous to the construction of a logical system codifying formal rules for deduction and interpretational rules for independence. This is strictly analogous to the precedence of good clinical practice standing before developed medical theory or to the precedence of good forensic argumentation standing before adequate legal theory. Moreover, contrary to the unfortunate way of presenting the issue in some logical textbooks, it seems dubious that knowledge of formal validity and invalidity arises from knowledge of logical forms. For, if knowledge of validity/invalidity of a given concrete argument were established by appealing to the known validity/invalidity of its abstract form the procedure would be flagrantly circular unless some kind of obscure or mystical knowledge of validity/invalidity of forms is postulated. This indicates the need for a piecemeal account of knowledge in practice of the validity and invalidity of concrete arguments composed of propositions and not of schemes. In order to fulfill this epistemic desideratum, next section postulates an information-theoretic viewpoint of implication and independence and an account of our corresponding judgments of validity and invalidity in practice.

One sense of 'information'

The tradition of looking at logical phenomena from an informational stance goes back as far as the XIX century. Logicians such as Boole, De Morgan, Jevons, and Venn already suggested that deducing is a sort of unpacking the information already contained in given premises. In the XX century this tradition is recovered by Carnap and Bar Hillel, Cohen and Nagel, and more recently by Corcoran. I

* This work is supported by the Spanish Ministry of Science and Technology under research projects HUM2006-04955/FISO and FF12008-00085.

follow Corcoran (1998 and 2007) in taking that the meaning of the word 'information' is given implicitly in the following conditions of his information-theoretic logic:

1. A premise-conclusion argument is valid if and only if the information in the conclusion is all contained in the premise-set. In other words, a premise-conclusion argument is valid if and only if the conclusion does not add any new information to the information already contained in the premises.
2. A premise-conclusion argument is invalid if and only if the information in the conclusion is not (all) contained in the premise-set.
3. Two propositions are logically equivalent if and only they have the same information. Sharing information content is necessary but not sufficient for the identity of propositions.
4. A tautology contains no information and a contradiction contains all the information pertinent to the given universe.
5. The negation of a given proposition contains the information not contained in the given proposition. It follows that a proposition and its negation do not have any information in common and that they jointly exhaust all the information pertain to their domain.
6. The information in a conjunction is the information of each of its conjuncts.
7. The information in a disjunction is the information shared by each of its disjuncts.
8. The information in a conditional is the information in the consequent that is not in the antecedent.
9. A universal proposition contains the information of each of its instances.
10. An existential proposition contains the information *shared* by each of its instances.

The previous informational clauses shape a logic which explains and systematized both the formal and the contentual traits of the relation of logical consequence between a set of propositions (the premises) and a single proposition (the conclusion). The identity criterion for propositions in this framework is very robust: Roughly, in order for "two" propositions to be just one proposition it is necessary and sufficient for them to pertain to the same informational domain, and to be composed of the very same logical and contentual concepts in the same ordered and number of occurrences. Notice that if two propositions do not pertain to the same informational domain they are bound to have different information content. Likewise, if two propositions are not composed by the same concepts in the same order and number of occurrences they may have different logical form or different content or neither. Moreover, information in this framework is said to be contained in propositions, not in isolated concepts.

Judging information-theoretic validity

For purposes of illustration it is assumed that all arithmetic examples in this paper make use of primitive concepts. Thus, "zero", "one", "two", "successor", "even", "square", etc., are all taken to be primitive concepts and the propositions considered all pertain to the domain of natural numbers.

a) -1 Three is prime

? If two is even then three is prime

This argument is valid. In fact, the conclusion contains less information than its premise because the information in the conditional of the conclusion is –according to condition eight above—, the information in "three is prime" which is not in "Two is even". By condition seven and condition five, this is the information shared by "three is prime" and "Two is not even". Hence we drop information in going from premise to conclusion.

b) -1 Three is prime

? If three is not prime then three is prime

This argument is also valid. Notice however that the conclusion drops no information from that contained in the premise. Since the antecedent of the conclusion is the negation of its consequent, by condition five they do not share any information. By condition seven, the information in "Three is prime" which is not in "Three is not prime" is exactly the information in the premise. Hence, the conclusion is logically equivalent to the premise.

c) -1 Three is prime

? If three is not prime then four is even

This argument is valid. One way of coming to judge validity is using again condition eight. The information in the conclusion is the information contained in "Four is even" which is not in "Three is not prime". By condition five, the information not contained in "Three is not prime" is exactly the information contained in "Three is prime". Thus the information shared by "Four is even" and "Three is prime" is less than the information contained in the premise "Three is prime. Hence, the argument is valid.

d) -1 Two is oblong and even or two is even and prime

? Two is even

The previous argument is valid. We judge it to be so by condition seven since the information in the conclusion "Two is even" is *all* the information shared by the disjuncts of the premise. Notice first, that the disjunction also implies the less informative "Some [number] is even" which is some of the information shared by both disjuncts. Notice also, that "Some [number] is even" also implies the less informative "Some [number] is oblong or even or prime".

e) -1 Zero is even and Two is even

? Every even [number] is even

This is also judged to be valid because the conclusion is tautological. By condition four, it lacks information and hence it does not add any information to the information already contained in the premise.

f) -1 One is the number that divides all and only the numbers that do not divide themselves

? One divides one

This argument is judged to be valid since the premise is contradictory. By condition four it contains all the information pertaining to the domain of natural numbers. Therefore, by condition one, it is valid.

g) -1 Two is not even

? Two is not both even and prime

This argument is valid. By condition five, the premise contains all the information not contained in "Two is even". By condition six "Two is even and prime" exceeds the information in "Two is even". Again by condition five, the negation of the conjunction is the information not contained in the conjunction and this is less than the information in the premise since the more conjuncts in the negated conjunction, the less information it contains.

h) ? It is not the case that two is even and two is not even

This argument is valid. By condition five the conclusion contains the information not contained in the proposition "Two is even and two is not even". Since a contradiction by condition four contains all the information, the information not in it is none. Hence by condition four, the conclusion is a tautology. By condition 1 the argument is valid.

Judging information-theoretic invalidity

The following examples illustrate judgment of non-containment between premises and conclusion.

a) -1 Every [number] is even or odd

? No [number] is even and odd

The argument is invalid because the conclusion adds information to the information already contained in the premise. Perhaps the first thing to notice is that the conclusion contains information and hence it is not tautological. In the second place, the information needed in the premise for the conclusion to follow is the one contained in "Every [number] is odd if and only if is not even" or some other equivalent proposition. Since this proposition is not in the premise-set, it is natural to conjecture that people wrongly judging validity of this argument are committing the fallacy of smuggling a premise.

b) -1 Two is prime

? Two is prime and two is even

This argument is obviously invalid. By condition six the information in the conjunction of the conclusion exceeds the information in the premise, which is its first conjunct. By condition two the argument is invalid.

c) ? Three is prime and three is not prime

This argument is judged to be invalid. By condition four, its conclusion exhausts all the pertinent information and in view of the fact that it lacks premises, by condition two, the conclusion clearly adds information.

d) -1 Two is not both odd and even

? Two is not odd

By condition five, the conclusion "Two is not odd" contains the information not contained in "Two is odd". Similarly, the premise contains the information not contained in "Two is both odd and even". By condition six the conjunction contains the information of each of its conjuncts. Therefore, the information not in the conjunction is less than the information not in one of its conjuncts. Thus the conclusion adds information to the premise and the argument is invalid.

e) -1 Zero is even

-2 One is even

-3 Two is even

……

…..

? Every [number] is even

This argument is also invalid. Notice first that it has an infinite premise-set. As a rule, the content of each of the singular instances does not logically imply the content of the corresponding universal closure. Presumably anyone judging this argument to be valid is smuggling a premise containing the information that there is no instance that is left out of the premise-set. None of the premises says that all of the instances have been considered. Hence by condition two the argument is invalid.

Conclusion

Logical methodology comprises the methods that are used to produce knowledge that a given conclusion is implicit in, or follows from, a given premise-set and also the methods that are used to produce knowledge that a given conclusion is not implicit in, or does not follow from, a given premise-set. This paper assumes an information-theoretic viewpoint in view of the fact that some important features of normal methodological practice remain hidden or perhaps just disregarded within the current paradigm of purely structural deductive systems and model-theoretic semantics. The rich and multifaceted experiences involved in our logical practice require for their intelligibility more than one concept of logical consequence. The main thesis of this paper is that information-theoretic logic is a complementary viable alternative.

References

Corcoran, J. (1998), 'Information-theoretic logic'. In Martínez, C., Rivas, U. and Villegas-Forero, L. (eds.), pp. 113-135.

Corcoran, J. (2007), 'Information-theoretic properties of truth-functional connectives', *The Bulletin of Symbolic Logic* 13, n. 3, p. 405.

Martínez, C., Rivas, U. and Villegas-Forero, L. (eds.) (1998), *Truth in perspective*, Aldershot, England and Brookfield, Vermont, Ashgate.

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, pp. 21-48.

Aproximación modal a la inferencia de nuevas teorías

Fernando Soler Toscano e Ignacio Hernández Antón
Universidad de Sevilla
fsoler@us.es / iha@us.es

Introducción

Los acercamientos lógicos al razonamiento abductivo proponen diversos modelos formales de lo que podemos llamar la *inferencia de la mejor explicación*. En resumen, si Θ representa una teoría (que puede entenderse como un conjunto de fórmulas de cierta lógica) y ϕ es una observación (una fórmula) que no se sigue de Θ (no es consecuencia lógica), la mejor explicación para ϕ en la teoría Θ será cierta fórmula α tal que ϕ sea consecuencia lógica de $\Theta \cup \{\alpha\}$. Los tratamientos formales del razonamiento abductivo (Aliseda, 2006) suelen imponer ciertos requisitos para que α pueda considerarse una buena explicación, o incluso la mejor explicación. En cualquier caso, estos acercamientos interpretan el razonamiento abductivo como un aumento en el conjunto de fórmulas de la teoría.

Ahora bien, la abducción lógica se queda corta si sólo puede dar cuenta de pequeños añadidos a las teorías (generalmente, además, sólo funciona bien en lógica proposicional, insuficientemente expresiva para representar teorías de cierta complejidad). De hecho, los tratamientos lógicos del razonamiento abductivo reciben constantemente la crítica de los filósofos de la ciencia, por no ser capaz de modelar modificaciones profundas en las teorías, como las que suponen los cambios de paradigma.

Hintikka (1998), en la *tesis de comprensión,* exige que el razonamiento abductivo incluya todas las operaciones por las que se generan nuevas teorías. Por ello, en este trabajo proponemos un acercamiento formal al razonamiento abductivo que permite no sólo la inferencia de nuevos hechos (abducción lógica tradicional) sino que puede explicar incluso modificaciones en la *lógica subyacente* a ciertas teorías. Pensemos en que cuando un cierto hecho ϕ no se sigue de la teoría Θ, ello se debe a que en la lógica que estamos usando dentro de nuestra teoría (habitualmente lógica clásica) ϕ no es consecuencia de Θ. Solemos resolverlo añadiendo nuevas fórmulas a Θ pero, ¿por qué no modificar nuestra lógica? Hay ciertos episodios bien conocidos por los historiadores de la ciencia, como el paso de la mecánica clásica a la mecánica cuántica, donde sería erróneo afirmar que todo se reduce a añadir nuevos postulados a la teoría, o

abandonar algunos de los antiguos. Cambia toda una concepción de la realidad, que conlleva una nueva forma de razonar con ella.

Modelos explicativos

Para tratar formalmente este tipo de evolución de teorías, realizamos un acercamiento basado en lógica modal, interpretando cada mundo como una *lógica posible*, es decir, una teoría, entendida ahora no sólo como un conjunto de postulados, sino que igualmente incluimos en ella las reglas que nos permitirán razonar con los datos, es decir, una *relación de consecuencia lógica* propia para cada mundo. Definimos una relación de accesibilidad $\Re$ entre mundos que nos indica cuándo, desde una lógica, podemos pasar a otra (cuándo desde una teoría podemos evolucionar a otra). Así, si desde una teoría w resulta accesible otra teoría w', es decir $w\Re w'$, entonces la teoría w podría evolucionar hasta w', modificándose tal vez no sólo el conjunto de proposiciones que conforman la teoría, sino igualmente su lógica subyacente, incluso cuestiones profundas como el carácter clásico o no clásico de las inferencias. Téngase en cuenta que las lógicas que caracterizan cada mundo podrán ser completamente diferentes. Esto significa que las condiciones para que cierta fórmula β sea consecuencia de un conjunto Γ de fórmulas dado, pueden diferir entre un mundo y otro. En un mundo podemos usar lógica clásica, en otro lógica no monótona, etc.

Formalmente, un *modelo explicativo* es una tupla:

$$M = \langle L, W, \Lambda, \Re, \pi \rangle$$

donde:

- L es un lenguaje formal, que tomamos como lenguaje base, común a todos los mundos.
- El conjunto W de mundos es no vacío y enumerable.
- Λ es un conjunto no vacío de lógicas, es decir, de conjuntos de 2^L.
- $\Re \subseteq W \times W$ es la relación de accesibilidad entre mundos.
- $\pi : W$ a Λ es una función que asigna a cada mundo $w \in W$ una lógica $\pi(w)$.

El lenguaje L_M del modelo explicativo funciona como un metalenguaje que se construye sobre el lenguaje base L, tomando todas las fórmulas de L y cerrándolas bajo la conjunción $\alpha \& \beta$, negación $\sim\alpha$ y los operadores modales (referidos a la relación de accesibilidad $\Re$, como se verá) $^{+}\alpha$ y $^{-}\alpha$. Es posible definir una disyunción $\alpha \nabla \beta$ e implicación $\alpha \Rightarrow \beta$ en términos de la negación y

conjunción, así como operadores modales duales $\Diamond^{+}\alpha$ y $\Diamond^{-}\alpha$ definidos tal como es habitual. Como es natural, ninguno de estos operadores de L_M ocurrirá en L. A cada fórmula $\varphi \in L_M$ asignamos un conjunto $\|\varphi\| \subseteq W$ de mundos que satisfacen φ. Escribimos, indistintamente, $w \in \|\varphi\|$ y $\langle M, w\rangle \models \varphi$ para indicar que el mundo w (que representa una teoría cuya lógica subyacente es $\pi(w) \in \Lambda$) perteneciente al modelo explicativo M satisface φ. El conjunto $\|\varphi\|$ se define inductivamente:

- $\|\varphi\| = \{w : \varphi \in \pi(w)\}$, si $\varphi \in L$
- $\|\alpha \,\&\, \beta\| = \|\alpha\| \cap \|\beta\|$
- $\| \sim\alpha\| = W - \|\alpha\|$
- $\|\Diamond^{+}\alpha\| = \Re^{-1}(\|\alpha\|)$
- $\|\Diamond^{-}\alpha\| = \Re(\|\alpha\|)$

Se puede observar que $\Box^{+}$ y $\Diamond^{+}$ son los operadores estándar modales interpretados en $\Re$, así como $\Box^{-}$ y $\Diamond^{-}$ son interpretados en $\Re^{-1}$.

Modalización del razonamiento abductivo

Ahora, un *problema abductivo* surge cuando en cierto mundo w tenemos una fórmula ϕ que no es verdadera (no es válida en la lógica de w o no se sigue de los hechos observados). Formalmente un problema abductivo es un par $\langle w, \phi\rangle$ tal que:

1. $\langle M, w\rangle \models \Diamond^{+}\phi$
2. $\langle M, w\rangle \models \Diamond^{+} \sim \phi$

La primera condición expresa que, mientras que ϕ no necesariamente es verdadera en w, existe una lógica accesible en la que sí lo es. La segunda condición indica que ϕ no es verdadera en todas las posibles evoluciones de la teoría w.

La *solución abductiva* a dicho problema viene dada por el paso a un mundo w' tal que $w \Re w'$ y ϕ sea verdadera en w' (es decir, que sea válida o se siga de

las observaciones). ¿Qué diferencia la nueva teoría w' de la antigua w? Se pueden haber añadido nuevos postulados (abducción clásica) o haberse cambiado el *estilo de razonamiento* (como ocurre en los cambios de paradigma). Por tanto, la fórmula $\alpha \in L$ será una solución abductiva a cierto problema abductivo $\langle w, \phi \rangle$ si y sólo si:

1. $\langle M, w \rangle \models \Box^{+}(\alpha \Rightarrow \phi)$
2. $\langle M, w \rangle \models \Diamond^{+}\alpha$
3. $\langle M, w \rangle \models \Diamond^{-}\Diamond^{+}(\alpha \,\&\, \sim\phi)$

La primera condición indica que α, junto con la teoría actual (la lógica de w) explica ϕ. La segunda condición dice que α es admisible en la teoría actual, y la última indica que α, por sí sola, es insuficiente para explicar ϕ, por tanto necesita la teoría actual. Estos tres criterios se corresponden con la *abducción consistente explicativa* que para Aliseda (2006) es la que posee mayor interés epistémico. Cada uno de los mundos w' a los que la segunda condición nos garantiza que podemos acceder es una teoría posible a la que podemos evolucionar desde w para explicar ϕ.

Caracterización de la abducción clásica

Veamos un ejemplo concreto de nuestro modelo que permite capturar el razonamiento abductivo en lógica clásica tal como ha sido estudiado en diversos formalismos. Sea L' el lenguaje de la lógica clásica de primer orden y $\models^{c}$ la relación de consecuencia lógica clásica. Podemos definir la *relación de consecuencia módulo* Θ (Makinson, 2003), donde $\Theta \subseteq L'$ como

$$\Gamma \models^{c}_{\Theta} \varphi \text{ si y sólo si } \Gamma, \Theta \models^{c} \varphi$$

Entonces, definimos el modelo:

$$M^{c} = \langle L', \Lambda, W, \Re, \pi \rangle$$

de forma que:

- W es el conjunto de todos los subconjuntos consistentes de L'
- $\Lambda = \left\{ \left\langle L', \models^{c}_{w} \right\rangle \mid w \in W \right\}$
- $\pi(w) = \left\langle L, \models^{c}_{w} \right\rangle$

- $\Re = \{\langle w, w' \rangle \in W^2 \mid w \subseteq w'\}$

Pues bien, M^c representa el conjunto de teorías clásicas, y los problemas abductivos en este modelo se corresponden con la noción de abducción en lógica clásica (Aliseda, 2006), pero expresada dentro de nuestro formalismo. Dado que la relación de accesibilidad $\Re$ es transitiva y reflexiva, se verifican los axiomas de S4, lo cual resulta de sumo interés para una caracterización estructural del razonamiento explicativo.

Conclusiones

Como se observa en el ejemplo de M^c, nuestro acercamiento permite introducir dentro del ámbito formal caracterizaciones del razonamiento abductivo que de otra forma sólo eran posibles en el nivel de la metalógica. Igualmente, al estudiarse el cambio epistémico al nivel de L_M, y no de L, queda abierta la puerta a estudios generales sobre la evolución de las teorías más allá de los habituales tratamientos del razonamiento abductivo centrados en una lógica particular. Ahora, la abducción ocurre en el terreno de L_M y la deducción en el de L. Ambos ámbitos tienen semánticas definidas separadamente. Uno de los objetivos de Hintikka (1998) era independizar el razonamiento abductivo del deductivo. Queda abierta la pregunta de hasta qué punto puede ser esto logrado siguiendo esta propuesta.

Referencias bibliográficas

Aliseda, A. (2006), *Abductive Reasoning: Logical Investigations into Discovery and Explanation*, Berlin, Springer.

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

Makinson, D. (2003), 'Bridges between classical and nonmonotonic logic', *Logic Journal of the IGPL* 11, pp. 69–96.

Relaciones de equivalencia: de máquinas de Turing a funciones parciales computables

José Pedro Úbeda Rives
Universitat de València
Jose.P.Ubeda@uv.es

1. Introducción

De las posibles realizaciones del modelo computacional "*máquinas de Turing*", Turing (1936) describe una caracterizada por tener una cinta infinita en ambas direcciones dividida en celdas, cada una de las cuales contiene un símbolo de un vocabulario $A = \{a_0, a_1, ..., a_n\}$, donde $n > 0$ y a_0 es el símbolo *blanco*. Además, tiene una *cabeza lectora/escritora* o *máquina* que puede estar en uno de los estados del conjunto $Q = \{q_0, ..., q_m\}$, donde q_0 es el estado *inicial*. La cabeza puede moverse una celda a la derecha (D), una celda a la izquierda (I) o permanecer quieta (P). Las acciones o *transiciones* que puede realizar la máquina se especifican por medio de un conjunto de instrucciones cada una de las cuales se describe por un quíntuplo de la forma

$$\langle q_i, a_j, q_k, a_r, m \rangle ,$$

donde $q_i, q_k \in Q$, $a_j, a_r \in A$ y $m \in \{D, I, P\}$. Dicho quíntuplo indica que, si la máquina está en el estado q_i leyendo el símbolo a_j, sustituye a_j por el símbolo a_r, hace el movimiento indicado por m y pasa al estado q_k. El conjunto de quíntuplos de una máquina ha de cumplir la condición de *consistencia*: no hay dos quíntuplos distintos con las dos primeras componentes iguales.

Cada máquina es una *descripción formal* de un algoritmo y puede identificarse con su conjunto consistente y finito de quíntuplos. El tipo de algoritmo que describe depende de *especificaciones* sobre los inputs y outputs. Así el algoritmo puede computar funciones parciales de n argumentos $(n > 0)$ o listar con o sin repetición el *grafo* de la función parcial que computan o el conjunto de inputs para los que la máquina se para o el conjunto de outputs que genera, etc.

Aunque sólo se consideren máquinas cuyos símbolos están incluidos en un vocabulario finito fijo A, hay tantas máquinas como conjuntos. Si además el conjunto de estados de cualquier máquina es un subconjunto de $\mathcal{Q} = \{q_0, q_1, \ldots, q_n, \ldots\}$, entonces sólo hay $\aleph_0$ máquinas. Puede demostrarse que estas son suficientes para describir cualquier algoritmo computado por una máquina.

Cuando las máquinas sobre un alfabeto A se codifican para ser datos de una máquina universal $\mathcal{U}$ para un tipo de algoritmo se convierten en *programas*, que denominaremos *programas de Turing*, que sirven junto con $\mathcal{U}$ para describir un algoritmo. Estos programas pueden *enumerarse* algoritmícamente:

$$P_0, P_1, \ldots, P_n, \ldots$$

dando lugar, para cada $m \in \mathbb{N}$ a una enumeración de todas las funciones parciales recursivas de m argumentos (si $m = 1$ no se usa el superíndice):

$$\varphi_0^m, \varphi_1^m, \ldots, \varphi_n^m, \ldots;$$

a una enumeración de los conjuntos de inputs para los que se paran las máquinas:

$$W_0, W_1, \ldots, W_n, \ldots;$$

a una enumeración de los conjuntos de outputs que generan las máquinas:

$$I_0, I_1, \ldots, I_n, \ldots;$$

y, para cada máquina universal $\mathcal{U}$ para listar, a las enumeraciones de listas con o sin repetición de grafos, de inputs y de outputs siguientes:

$$\mathcal{U}_0^{rg}, \mathcal{U}_1^{rg}, \ldots, \mathcal{U}_n^{rg}, \ldots; \qquad \mathcal{U}_0^{g}, \mathcal{U}_1^{g}, \ldots, \mathcal{U}_n^{g}, \ldots;$$

$$\mathcal{U}_0^{ri}, \mathcal{U}_1^{ri}, \ldots, \mathcal{U}_n^{ri}, \ldots; \qquad \mathcal{U}_0^{i}, \mathcal{U}_1^{i}, \ldots, \mathcal{U}_n^{i}, \ldots;$$

$$\mathcal{U}_0^{ro}, \mathcal{U}_1^{ro}, \ldots, \mathcal{U}_n^{ro}, \ldots; \qquad \mathcal{U}_0^{o}, \mathcal{U}_1^{o}, \ldots, \mathcal{U}_n^{o}, \ldots.$$

Los tipos de algoritmos descritos por los programas de Turing dan lugar a relaciones de equivalencia entre programas. Sean P_a y P_b dos programas, entonces podemos definir las siguientes relaciones de equivalencia:

$$P_a \equiv_{\varphi} P_b \Leftrightarrow \varphi_a \simeq \varphi_b,$$

$$P_a \equiv_{W} P_b \Leftrightarrow W_a = W_b,$$

$$P_a \equiv_{I} P_b \Leftrightarrow I_a = I_b,$$

$$P_a \equiv_{\mathcal{U}^{\beta}} P_b \Leftrightarrow \mathcal{U}_a^{\beta} = \mathcal{U}_b^{\beta},$$

donde $\beta \in \{rg, g, ri, i, ro, o\}$. Las relaciones de equivalencia entre programas, dada una enumeración de programas, pueden identificarse con relaciones de equivalencia entre sus índices, de forma que

$$a \equiv b \Leftrightarrow P_a \equiv P_b.$$

Bajo este supuesto se habla de relaciones de equivalencia entre programas que son recursivas, recursivamente enumerables, etc.

En esta comunicación se analizan nueve clases de relaciones de equivalencia entre programas: $\mathcal{R}_{\varphi}$, $\mathcal{R}_{W}$, $\mathcal{R}_{I}$, $\mathcal{R}_{\mathcal{U}^{rg}}$, $\mathcal{R}_{\mathcal{U}^{g}}$, $\mathcal{R}_{\mathcal{U}^{ri}}$, $\mathcal{R}_{\mathcal{U}^{i}}$, $\mathcal{R}_{\mathcal{U}^{ro}}$ y $\mathcal{R}_{\mathcal{U}^{o}}$. Cada clase $\mathcal{R}_{\alpha}$, donde $\alpha \in \{\varphi, W, I, \mathcal{U}^{rg}, \mathcal{U}^{g}, \mathcal{U}^{ri}, \mathcal{U}^{i}, \mathcal{U}^{ro}, \mathcal{U}^{o}\}$, consta de las relaciones de equivalencia $\equiv$ tales que, para cualesquiera dos programas P_a y P_b, se cumple

$$P_a \equiv P_b \rightarrow P_a \equiv_{\alpha} P_b.$$

2. Máquinas universales para listar

Utilizando la tesis de Church vamos a describir por medio de un programa seis series de máquinas de Turing que aplicadas a un programa de Turing P_x producen una lista. Cada una de las seis series consta de infinitas máquinas de Turing para listar, una para cada función recursiva f. Para cada función recursiva f la primera y segunda serie producen, respectivamente, las máquinas ${}_f\mathcal{U}^{rg}$ y ${}_f\mathcal{U}^{g}$ que listan, con o sin repetición, el grafo de la función φ_x computada por P_x, la tercera y cuarta producen, respectivamente, las máquinas ${}_f\mathcal{U}^{ri}$ y ${}_f\mathcal{U}^{i}$ que listan el dominio de φ_x y la quinta y sexta producen las máquinas ${}_f\mathcal{U}^{ro}$ y ${}_f\mathcal{U}^{o}$ que listan el rango de φ_x.

Para describir el programa usaremos la función

$$P_x^t(s) = \begin{cases} \varphi_x(s), & \text{si } \varphi_x(s)\downarrow \text{ en } t \text{ o menos pasos;} \\ \uparrow, & \text{en otro caso;} \end{cases}$$

que es parcial recursiva y por la Tesis de Church-Turing hay una máquina $\mathcal{U}$ que aplicada a un programa de Turing P, un número t y una palabra $s \in A^*$, imita t pasos de la computación de la máquina con programa P aplicada al input s. Usaremos también la notación $C^\frown B$ para denotar la lista obtenida al concatenar la lista C y la B en ese orden. La lista con componentes $a, b, c, \ldots$ la denotamos por $[a, b, c, \ldots]$ y la lista vacía por $[]$. La máquina descrita por el programa no se para nunca y va dando como outputs los distintos elementos de la lista. Cada máquina tiene realmente tres inputs: el primero es la función recursiva f o un índice suyo (este input no aparece explícito en el programa y se supone que se tiene un subprograma que la calcula), el segundo es el índice x del programa de Turing P_x y el tercero es un número perteneciente a $\{0, 1, 2, 3, 4, 5\}$ que indica si los elementos listados pertenezcan al grafo (el número 0 y 1), al dominio (el 2 y 3) o al rango (el 4 y 5) de la función φ_x.

PROGRAMA

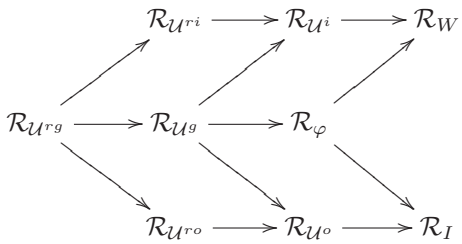
principal
  input x % La función φ_x de la que se quiere una lista
  input m % 0 para ${}_f\mathcal{U}^{rg}$, $1 \Rightarrow_f \mathcal{U}^{g}$, $2 \Rightarrow_f \mathcal{U}^{ri}$, $3 \Rightarrow_f \mathcal{U}^{i}$, $4 \Rightarrow_f \mathcal{U}^{ro}$ y $5 \Rightarrow_f \mathcal{U}^{o}$.
  $lista = \emptyset$
  $pasos = 0$ % índice para contar los pasos en la computación
  $entrada = 0$ % índice para los inputs de φ_x
  while true do
    $pasos = pasos + 1 + f(entrada)$
    $j = 0$
    while $j \leq entrada$ **do**
      if $P_x^{pasos}(j)\downarrow$ **then**
        $z = P_x^{pasos}(j)$
        hace$(m, lista, z, j)$ % subprograma
      $j = j + 1$
    endwhile
    $entrada = entrada + 1$

```
    endwhile
hace(m, lista, z, j) % SUBPROGRAMA
    if m = 0 or m = 1 then % Listar grafo(φ_x)
        if m = 0 or ⟨j, z⟩ ∉ lista then
            lista = lista⌢[⟨j, z⟩]
            output ⟨j, z⟩
    if m = 2 or m = 3 then % Listar W_x
        if m = 2 or j ∉ lista then
            lista = lista⌢[j]
            output j
    if m = 4 or m = 5 then % Listar I_x
        if m = 4 or z ∉ lista then
            lista = lista⌢[z]
            output z
```

3. Propiedades de las clases $\mathcal{R}_\alpha$

La relación de identidad entre programas pertenece a cada clase $\mathcal{R}_\alpha$ y determina $\aleph_0$ clases de equivalencia unitarias. Además, cada relación $\equiv_\alpha$ determina $\aleph_0$ clases de equivalencia cada una de las cuales consta de $\aleph_0$ programas. A partir de estos dos hechos se demuestra que cada clase $\mathcal{R}_\alpha$ consta de $2^{\aleph_0}$ relaciones.

Las relaciones de inclusión entre dichas clases pueden resumirse en el siguiente diagrama, donde la flecha es la inclusión propia y $\mathcal{U}$ es ${}_f\mathcal{U}$ para una función recursiva dada. Si f y g son dos funciones recursivas distintas no pueden establecerse en general relaciones de inclusión entre $\mathcal{R}_{{}_f\mathcal{U}^\beta}$ y $\mathcal{R}_{{}_g\mathcal{U}^\beta}$.



Para establecer la inclusión propia entre $\mathcal{R}_\varphi$ y $\mathcal{R}_W$, entre $\mathcal{R}_{\mathcal{U}^{rg}}$ y $\mathcal{R}_{\mathcal{U}^{ri}}$ y entre $\mathcal{R}_{\mathcal{U}^{g}}$ y $\mathcal{R}_{\mathcal{U}^{i}}$ es suficiente considerar que si φ es parcial recursiva, la función

$$\chi(x) = \begin{cases} \varphi(x) + 1, & \text{si } \varphi(x) \downarrow; \\ \uparrow, & \text{en otro caso} \end{cases}$$

es tal que $Dom(\varphi) = Dom(\chi)$, pero no se tiene $\varphi \simeq \chi$, excepto si $\varphi = \emptyset$, la función no definida en ningún sitio. Para la inclusión propia entre $\mathcal{R}_\varphi$ y $\mathcal{R}_I$, entre $\mathcal{R}_{\mathcal{U}^{rg}}$ y $\mathcal{R}_{\mathcal{U}^{ro}}$ y entre $\mathcal{R}_{\mathcal{U}^{g}}$ y $\mathcal{R}_{\mathcal{U}^{o}}$ es suficiente considerar la función ψ tal que $\psi(0) \uparrow$ y

$$\psi(x') = \begin{cases} \varphi(x), & \text{si } \varphi(x) \downarrow; \\ \uparrow, & \text{en otro caso} \end{cases}$$

ya que $Rg(\varphi) = Rg(\psi)$ y no se tiene $\varphi \simeq \psi$, excepto si $\varphi = \emptyset$.

Cada clase $\mathcal{R}_\alpha$ es un retículo con respecto a las operaciones $\wedge$ y $\vee$ tales que, para todo $r, s \in \mathcal{R}_\alpha$, $r \wedge s = r \cap s$ y $r \vee s$ es la menor relación de equivalencia t tal que $r \subseteq t$ y $s \subseteq t$. Esto se sigue al estar cerradas las $\mathcal{R}_\alpha$ con respecto a $\wedge$ y $\vee$. Además, cada $\mathcal{R}_\alpha$ tiene la identidad $=$ que es recursiva como elemento mínimo y $\equiv_\alpha$ que no es recursivamente enumerable como elemento máximo.

Las relaciones de equivalencia recursivas no forman un retículo (véase el Apéndice A), pero las relaciones de equivalencia recursivamente enumerables si forman un retículo, ya que, si s y t son relaciones de equivalencia recursivamente enumerables, entonces también son recursivamente enumerables $s \vee t$ y $s \wedge t$ (véase el Apéndice B). De esto se sigue que hay un subretículo $\mathcal{R}_{re} \subset \mathcal{R}_{\mathcal{U}^{rg}}$ que consta de relaciones de equivalencia recursivamente enumerables. Este retículo posee como elemento mínimo la igualdad, pero no tiene elemento máximo.

Si se define un algoritmo como una clase de equivalencia, como proponen Yanofsky (2006), Shore en Buss (2001) y Moschovakis (2001), ¿a cuál de las $2^{\aleph_0}$ relaciones de equivalencia hemos de referirnos? Por supuesto si se desea que la igualdad entre algoritmos sea decidible, dicha relación debe ser recursiva, pero existen $\aleph_0$ relaciones de equivalencia recursivas. Si lo único que se desea es poder listar todos los programas que describen el mismo argumento, dicha relación tiene que ser recursivamente enumerable, pero existen $\aleph_0$ relaciones de equivalencia recursivamente enumerables.[1]

A. Las relaciones de equivalencia recursivas no forman un retículo

Sea $\langle a, b\rangle$ una función de *parificación* biyectiva de $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ y π_1 y π_2 sus funciones inversas tales que

$$n = \langle \pi_1(n), \pi_2(n)\rangle\,.$$

Supongamos dada una codificación de las configuraciones para máquinas de Turing y de sucesiones de configuraciones tal que, si t es la codificación de la sucesión de configuraciones $c(t)$ de la forma

$$u_1 q_{i_1} v_1, u_2 q_{i_2} v_2, \cdots, u_s q_{i_s} v_s,$$

entonces es recursiva la función

$$f(e,t) = \begin{cases} 1, & \text{si } P_e \text{ puede recorrer la sucesión } c(t) \text{ de configuraciones} \\ 0, & \text{en otro caso} \end{cases}$$

Sea $\equiv_c$ la relación de equivalencia tal que, para todo $n, m \in \mathbb{N}$, $n \equiv_c m$ si y sólo si $n = m$ o cumple las siguientes condiciones

1. $\pi_1(n) = \pi_1(m)$,
2. $f(\pi_1(n), \pi_2(n)) = f(\pi_1(m), \pi_2(m)) = 1$ y

[1] Blass, Dershowitz y Gurevich (2008) defienden que es imposible definir algoritmo por medio de relaciones de equivalencia.

3. la primera configuración de $\pi_2(n)$ es inicial e igual a la primera de $\pi_2(m)$, que también es inicial.

es decir, $\langle e,t\rangle$ es equivalente a $\langle d,s\rangle$ si son iguales o $e=d$ y t y s son codigos de las sucesiones t' y s' de configuraciones de e tal que t' es una subsucesión inicial de s' o s' es una subsucesión inicial de t'. La relación $\equiv_c$ es recursiva. Igualmente lo es la siguiente relación de equivalencia $\equiv_f$ tal que, para todo $n,m\in\mathbb{N}$,

$$n \equiv_f m$$

si y sólo si $n=m$ o $\pi_1(n)=\pi_1(m)$, $f(\pi_1(n),\pi_2(n))=f((\pi_1(m),\pi_2(m))=1$ y la última configuración de $\pi_2(n)$ y la última de $\pi_2(m)$ son finales; es decir, $\langle e,t\rangle \equiv_f \langle d,s\rangle$ si y sólo si o $\langle e,t\rangle=\langle d,s\rangle$ o t y s son sucesiones de configuraciones para la máquina e cuyo último término es una configuración final.

Es fácil demostrar que si $(\equiv_f \vee \equiv_c)$ es recursiva, entonces el problema de la parada es resoluble. Supongamos que $(\equiv_f \vee \equiv_c)$ es recursiva. Entonces $P_e(e)\downarrow$ si y sólo si

$$\langle e,t\rangle\,(\equiv_f \vee \equiv_c)\,\langle e,s\rangle$$

donde t es la configuración inicial para el input e y s es una configuración final para e.

B. Relaciones de equivalencia recursivamente enumerables

Sea f una función recursiva tal que $f(i)=j$ es el índice de una relación de equivalencia S recursivamente enumerable, es decir,

$$xSy \text{ si y sólo si } \langle x,y\rangle \in W_j.$$

Vamos a mostrar que $\bigvee_{i\in\mathbb{N}} W_{f(i)}$ es recursivamente enumerable. Para ello presentamos un programa en pseudocódigo, que por la Tesis de Church nos dará el resultado buscado. Por supuesto si no tenemos una enumeración o una función f recursiva como la supuesta arriba, no podemos establecer para un conjunto $\mathbf{C}$ que no es recursivamente enumerable que $\bigvee_{i\in\mathbf{C}} W_i$ sea una relación de equivalencia recursivamente enumerable.

En el programa usamos para las listas las funciones *cabeza*, *cuerpo*, $\frown$, *ultimo* (el último elemento de la lista) y $\notin$ (no ocurre en la lista).

PROGRAMA

input x,y % Contestará **si** si $x\left(\bigvee_{i\in\mathbb{N}} R_{f(i)}\right)y$, no parará en otro caso
  if $x=y$ **then**
    output si
    end
  $seguir$ = **true**
  $pasos = 0$
  $entrada = 0$

$lista\_numeros = \emptyset$
$lista\_pares = \emptyset$
$tiras = \emptyset$
while *seguir* **do**
  $pasos = pasos + 1$
  $j = 0$ % Índice para los argumentos
  while $j \leq entrada$ **do**
    $k = 0$ % Índice para las relaciones
    while $k \leq entrada$ **do**
      if $P^{pasos}_{f(k)}(j) \downarrow$ **and** $j \notin lista\_numeros$ **then**
        $lista\_numeros = lista\_numeros^\frown[j]$
        if $\pi_1(j) \neq \pi_2(j)$ **then**
          $lista\_pares = lista\_pares^\frown[[\pi_1(j), \pi_2(j)]]$
          $tiras1 = tiras$
          $tiras = tiras^\frown[[\pi_1(j), \pi_2(j)], [\pi_2(j), \pi_1(j)]]$
          while $tira1 \neq \emptyset$ **do**
            $t = cabeza(tira1)$
            $tira1 = cuerpo(tira1)$
            if $cabeza(t) = \pi_2(j)$ **and** $\pi_1(j) \notin t$ **then**
              $tira1 = tira^\frown[[\pi_1(j)]^\frown t]$
              if $\pi_1(j) = x$ **and** $ultimo(t) = y$ **then**
                output si
                end
            if $cabeza(t) = \pi_1(j)$ **and** $\pi_2(j) \notin t$ **then**
              $tira1 = tira^\frown[[\pi_2(j)]^\frown t]$
              if $\pi_2(j) = x$ **and** $ultimo(t) = y$ **then**
                output si
                end
            if $ultimo(t) = \pi_2(j)$ **and** $\pi_1(j) \notin t$ **then**
              $tira1 = tira^\frown[t^\frown[\pi_1(j)]]$
              if $cabeza(t) = x$ **and** $\pi_1(j) = y$ **then**
                output si
                end
            if $ultimo(t) = \pi_1(j)$ **and** $\pi_2(j) \notin t$ **then**
              $tira1 = tira^\frown[t^\frown[\pi_2(j)]]$
              if $cabeza(t) = x$ **and** $\pi_2(j) = y$ **then**
                output si
                end
          endwhile
      $k = k + 1$
    endwhile
    $j = j + 1$
  endwhile
  $entrada = entrada + 1$
endwhile

Referencias bibliográficas

Blass, A., Dershowitz, N. y Gurevich, Y. (2008), 'When are two algorithms the same?', *Technical Report MSR-TR-2008*, Microsoft, February 2008. Accesible en https://research.microsoft.com/en-us/um /people/gurevich.

Buss, S., Kechris, A., Pillay, A. y Shore, R. (2001), 'The prospects for mathematical logic in the twenty-first century', *Bulletin of Symbolic Logic* 7, pp. 169–196.

Davis, M. (1965), *The undecidable*, New York, Raven Press.

Moschovakis, Y. N. (2001), 'What is an algorithm?', en Engquist, B. y Schmid, W. (eds), *Mathematics Unlimited*, Berlin, Springer, pp. 919–936.

Turing, A. M. (1936), 'On computable numbers, with an application to the entscheidungproblem', *Proceedings of London Mathematical Society* 42, pp. 230–265. Corrección, *ibidem*, 43, pp. 544–546. Reimpreso en Davis (1965), pp. 155–222.

Yanofsky, N. Y. (2006), 'Towards a definition of an algorithm', *arXiv: math.LO/0602053v1*, pp. 1–38.

Modelos formales para la combinación de creencias en conflicto

Julián Velarde Lombraña
Universidad de Oviedo
velarde@uniovi.es

Introducción

La teoría de la evidencia, iniciada por Dempster en los años 60 del pasado siglo (1967), desarrollada luego por Shafer y conocida generalmente como la teoría de Dempster-Shafer (TDS), proporciona un modelo formal para computar las funciones de creencia simples y la combinación de funciones, lo que permite cuantificar la amplitud de evidencia que conlleva un mensaje (información) complejo y no susceptible de descomposición en componentes independientes. El problema central con el que se encuentra esta teoría es la combinación de creencias en conflicto. Zadeh (1984) aduce un ejemplo en el que la regla de combinación de Dempster produce resultados considerados usualmente insatisfactorios o contra-intuitivos. Desde entonces, muchos autores han utilizado el ejemplo de Zadeh, bien para criticar la TDS en su conjunto, bien para proponer reglas de combinación alternativas que eviten resultados anómalos cuando se presenta evidencia muy conflictiva.

Elementos de la teoría de la evidencia

La teoría de Dempster-Shafer (TDS) está normalmente basada sobre un conjunto Θ no vacío, exclusivo, exhaustivo y supuestamente finito de hipótesis, estados o acontecimientos: $\Theta = \{H_1 , H_2, \ldots, H_n\}$. A partir de Θ se obtiene el conjunto 2^Θ, llamado *marco de discernimiento*, que comprende las 2^n proposiciones (estados, hipótesis) A de Θ:

$$2^\Theta = \{A / A \subseteq \Theta\} = \{\varnothing, \{H_1\}, \{H_2\}, \ldots, \{H_1, H_2\}, \ldots, \Theta\}$$

A cada una de estas proposiciones A de 2^Θ se asigna una distribución de masa de evidencia *m(A)*, asociada a la creencia y la plausibilidad. La función de masa *m(A)* representa la confianza prestada estrictamente a A (en qué medida la evidencia soporta la hipótesis A).

Dado un marco de discernimiento 2^Θ, una función $m : 2^\Theta \rightarrow [0, 1]$ es llamada una *función de masa*, si y sólo si,

1. $m(\varnothing) = 0$
2. $\Sigma_{a \subseteq \Theta} m(A) = 1$

Los subconjuntos $A \subseteq \Theta$ tales que $m(A) > 0$ son llamados *elementos focales*, y el par $\langle F,m \rangle$, en donde F denota el conjunto de todos los elementos focales inducidos por m es llamado *cuerpo de evidencia.*

A partir de las funciones de masa es definida la función de *credibilidad*. La credibilidad de un conjunto de elementos A en un marco de discernimiento 2^{Θ} es definida como la credibilidad total de A: la suma de todas las masas asignadas a los elementos contenidos en A más la masa atribuida a la propia A:

$$Cred(A) = \Sigma_{a \subseteq \Theta} m\,(X)$$

Regla de combinación de funciones de credibilidad (regla de Dempster)

La regla de combinación de Dempster permite calcular, a partir de dos funciones de credibilidad sobre el mismo marco de discernimiento 2^{Θ}, una nueva función de credibilidad $\oplus$, llamada su *suma ortogonal*, basada sobre la evidencia combinada. Dadas, sobre el mismo marco de discernimiento, m_1 con A_i elementos focales y m_2 con B_j elementos focales, la nueva evidencia que apoya a C es la suma de la evidencia de todas las intersecciones entre los conjuntos A_i y B_j que den como resultado C. Así:

$$[m_1 \oplus m_2]\,(C) = \Sigma_{ai \cap Bj = c} m_1\,(A_i) \cdot m_2\,(B_i)$$

El problema de este esquema es que puede atribuir una masa de evidencia no nula al conjunto vacío, lo que contradice las reglas de base para las funciones de credibilidad; la masa $m(\varnothing)$ asociada al conjunto vacío refleja el conflicto entre las dos fuentes de evidencia. Para resolver este problema es necesario descartar los productos $m_1(A_i) \,.\, m_2(B_j)$ tales que $A_i \cap B_j = \varnothing$, redistribuyendo $m(\varnothing) = K$ entre los restantes productos $m_1(A_i) \,.\, m_2(B_j)$ del marco de discernimiento tales que $A_i \cap B_j \neq \varnothing$; esto es, cada uno de los otros valores-masa es dividido por $1 - K$ (de manera equivalente: multiplicado por $(1 - K)^{-1}$), siendo

$$K = \Sigma_{ai \cap Bj = \phi} m_1\,(A_i) \cdot m_2\,(B_i) > 0$$

Tales subconjuntos quedan así *normalizados*. La regla de combinación de Dempster deviene con la normalización:

$$[m_1 \oplus m_2]\,(C) = \Sigma_{Ai \cap Bj = C} \frac{m_1\,(A_i) \cdot m_2\,(B_j)}{1 - K}$$

El factor de normalización, por una parte, produce convergencia hacia la opinión dominante; y por otra parte tiene el efecto de ignorar completamente el conflicto, atribuyendo cualquier masa de evidencia en conflicto al conjunto nulo. Zadeh (1984), en su reseña del libro fundamental de Shafer (1976), aduce un ejemplo en el que la regla de combinación de Dempster produce resultados considerados usualmente insatisfactorios o contra-intuitivos. Muchos autores han utilizado este ejemplo, bien para criticar la teoría de Dempster-Shafer en su conjunto, bien para proponer reglas de combinación que eviten estos resultados contra-intuitivos.

Los modelos posibilistas para la combinación de creencias (evidencia)

Dentro del esquema posibilista (Dubois y Prade, 1988, 1992, 2000) la información obtenida de la fuente (experto) *i* viene representada por la función de distribución de posibilidad π (análoga a la función de masa *m* en la teoría de la evidencia) sobre un universo de conocimiento U en el intervalo [0,1]

$$\pi : U \rightarrow [0,1]$$

Donde π(u) es interpretada como el grado de posibilidad de que *u* sea verdadera. Se dice que π está *normalizada*, si existe un valor $u \in U$ tal que $\pi(u) = 1$. π(u) = 0 significa que *u* es imposible; π(u) = 1 significa que *u* es completamente satisfecha (posible); π(u) > π(u') significa que *u* es preferida a (es más plausible que) *u'*.

Por analogía con la medida de probabilidad P de los todos probabilistas, introduce Zadeh (1978) en los todos posibilistas una medida de posibilidad Π, como una aplicación de los subconjuntos A de U en el intervalo [0,1]

$$\Pi : 2^U \rightarrow [0,1]$$

La medida de posibilidad de los subconjuntos $A_i \in U$ viene inducida (determinada unívocamente) mediante una función de distribución de posibilidad π que viene dada por la fórmula:

$$\Pi(A) =_{def} \max\{\pi(u) / u \in A\}$$

Y tal que:

1. $\Pi(\varnothing) = 0$
2. $0 \leq \Pi(A) \leq 1 \quad \forall A \in U$
3. Si $A \subseteq B \rightarrow \Pi(A) \leq \Pi(B) \; \forall A, B \in U$

En el marco de la teoría de la posibilidad han sido propuestas varias reglas de combinación de evidencia que toman en cuenta de manera explícita la información contextual sobre las fuentes: su grado de fiabilidad y el grado de conflicto entre ellas.

Hay básicamente dos modos extremos de combinación simétricos: el modo *conjuntivo*, cuando todas las fuentes concuerdan y son igualmente fiables; y el modo *disyuntivo*, cuando las fuentes no concuerdan y se sabe con seguridad que *al menos una* es fiable, pero no se sabe cuál.

Sea π_i la distribución de posibilidad proporcionada por la fuente *i* (*i* = 1, …, n); tenemos, entonces:

$$\text{Regla conjuntiva: } \pi_{\cap}(u) = \prod_{i=1}^{n} \pi_i(u) \qquad \forall u \in U$$

$$\text{Regla disyuntiva: } \pi_{\cup}(u) = \coprod_{i=1}^{n} \pi_i(u) \qquad \forall u \in U$$

En donde $\cap$ y $\cup$ son dos operaciones de $[0,1] \times [0,1]$ sobre [0,1] que cumplen la relación de dualidad: $a \cup b = 1-(1-a) \cap (1-b)$ (expresión de las leyes de De Morgan).

Posibles operaciones de este tipo, que utilizan los modelos posibilistas, son las llamadas normas y conormas triangulares (t-normas y t-conormas). Hay una amplia variedad de instancias de t-normas y t-conormas. Una de ellas es la propuesta por Zadeh en el ámbito de la teoría de lo difuso (para la t-norma T y la t-conorma S): T(x , y) = min (x , y); y S(x , y) = max (x , y).

En la combinación de informaciones, el modo más utilizado es el conjuntivo; en él la operación *min* corresponde a una consideración puramente lógica del proceso de combinación. La fuente que asigna el *menor* grado de posibilidad a un valor dado es considerada como *la mejor* informada respecto de ese valor. Con *min*, cuando todas las fuentes están en completo acuerdo, no hay efecto refuerzo. En el modo conjuntivo, el resultado puede no quedar normalizado, i. e., cabe la posibilidad de que $\neg\exists$ u, $\pi\cap$ (u) = 1, lo que expresa que hay conflicto entre las fuentes. De aquí cabe inferir que el modo conjuntivo tiene sentido si todas las fuentes π_i tienen un solapamiento significativo, i. e., $\exists$u, $\forall$i, π_i = 1, lo que expresa que hay al menos un valor *u* que todas las fuentes consideran completamente posible. Pero si $\forall$u, $\pi\cap$ (u) es sensiblemente inferior a 1, entonces queda patente un alto grado de conflicto, y por ello puede resultar más adecuado el modo de combinación disyuntivo.

Finalmente, en caso de conflicto parcial, puede resultar adecuado normalizar la regla conjuntiva con vistas a cumplir el requisito $\max_{u\in U}\ \pi(u) = 1$, obteniendo así una nueva regla de combinación, dentro del modelo posibilista, análoga a la regla de Dempster, dentro del modelo de la teoría de la evidencia:

$$\pi(u) = \frac{\pi\cap(u)}{d\,(\pi_1,\ \pi_2)} \quad \forall u \in U$$

En donde $d(\pi_1,\ \pi_2)$ es la amplitud de la intersección de π_1 y π_2, llamada índice de consistencia, es una medida de solapamiento entre dos distribuciones de posibilidad, definida por

$$d(\pi_1,\ \pi_2) = \text{Sup.}_{u\in U} \min\{\pi_1(u),\ \pi_2(u)\} = \text{Sup.}\ \pi\cap$$

$1 - d(\pi_1,\ \pi_2)$ representa, por tanto, la medida global de conflicto entre las distribuciones de posibilidad π_1 y π_2. La normalización borra el conflicto entre las fuentes, pudiendo generar problemas en aquellos casos en que hay muy amplio conflicto. Si se quiere seguir teniendo en cuenta el conflicto parcial entre las fuentes se puede buscar una fórmula para descontar del resultado el peso correspondiente a la falta de normalización. Una fórmula posible (que corresponde a una interpretación de los grados de posibilidad ordenados) es la siguiente:

$$\forall u \in U$$

$$\pi'(u) = \max\{\pi(u), 1 - d(\pi_1,\pi_2)\} = \max\left[\frac{\min\{\pi_1(u),\pi_2(u)\}}{d(\pi_1,\pi_2)}, 1 - d(\pi_1,\pi_2)\right]$$

El modelo posibilista proporciona también una regla de combinación que toma en cuenta de manera explícita el conocimiento disponible acerca del valor de los coeficientes de fiabilidad de las fuentes. A este respecto se presentan varias posibilidades (Dubois y Prade, 1992). Una de ellas consiste en incluir la fiabilidad de cada fuente *antes* de la combinación, para compensar su diferente fiabilidad y hacerlas así totalmente (al menos igualmente) fiables antes de la combinación. Así, si se conoce el grado de certeza R de que una fuente *i* es fiable, entonces cabe aplicar la siguiente regla de ajuste que cambia la distribución de posibilidad π_i , proporcionada por la fuente *i*, en π'_i:

$$\pi'_i = \max (\pi_i, 1 - R_i)$$

En donde R_i es el grado de certeza de que la fuente *i* es fiable. Si $R_i = 1$ (esto es, la fuente *i* es completamente fiable), entonces $\pi'_i = \pi_i$. Cuando $R_i = 0$ (esto es, la fuente *i* no es nada fiable), entonces $\pi'_i = 1$, lo que significa ignorancia total. Las fuentes, una vez reajustadas de acuerdo con sus niveles de fiabilidad, son tomadas como absolutamente fiables y son combinadas mediante la regla conjuntiva. El inconveniente de esta regla es que cuando $R_i = 0$, entonces el soporte de la distribución reajustada deviene el universo entero.

Conclusiones

Cuando se presenta evidencia conflictiva, la regla de Dempster para la combinación de creencias genera resultados que no reflejan adecuadamente la actual distribución de creencias. La teoría de la posibilidad permite modelar varias reglas alternativas que toman en cuenta el tipo de conflicto entre las creencias (o fuentes de evidencia) y la *fiabilidad* de las mismas.

Referencias bibliográficas

Dempster, A. (1967), 'Upper and lower probabilities induced by a multivalued mapping', *Annals of Mathematical Statistics* 38, pp. 325-339.

Dubois, D. y Prade, H. (1988), 'Representation and combination of uncertainty with belief functions and possibility measures', *Computational Intelligence* 4, pp. 244-264.

— (1992), 'Combination of fuzzy information in the framework of possibility theory', en Abidi, M. y González, R. (eds.), *Data Fusion in Robotics and Machine Intelligence*, Boston, Academic Press, pp. 481-505.

— (2000), 'Possibility theory in information fusion', *Proceedings of the Third International Conference on Information Fusion*, Paris, 1-2, pp. 6-19.

Shafer, G. (1976), *A Mathematical Theory of Evidence*, Princeton, N. J., Princeton University Press.

Zadeh, L. A. (1978), 'Fuzzy sets as a basis for a theory of possibility', *Fuzzy Sets and Systems* 1, pp. 3-28.

— (1988), 'Review of Mathematical Theory of Evidence, by G. Shafer', *AI Magazine* 5, pp. 81-83.

Sección B

Filosofía del lenguaje, filosofía de la mente, epistemología

Against original intentionality

Marc Artiga Galindo
LOGOS – Universitat de Girona
Skartiga@hotmail.com

It is generally assumed that intentionality is exhibited, at least, by mind and language. This claim is often followed by another one: the intentionality of language is derived from the intentionality of mind. The aim of this paper is to show that the arguments underpinning this assumption do not withstand careful examination, and hence that the thesis of Derived Intentionality (DI) should be abandoned:

(DI) The intentionality of language derives from the intentionality of mind.

DI has been rarely argued for as a general principle. If we look at the literature, we see that the discussion is not cashed out in terms of intentionality but rather in terms of meaning. How can a theory about meaning determine whether DI is true?

There are two relevant points to be made here. First, it important to notice that in this debate to say that language exhibits intentionality amounts to saying that language has meaning. In the present discussion we are interested in the fact that linguistic expressions essentially refer to other things, they are *about* something and this is just to wonder about the fact that language has meaning.

Secondly, it has traditionally been held that meaning has to aspects: speaker and linguistic meaning. Speaker meaning is what speakers intend to express by a sign, what a speaker *means* by it. On the other hand, linguistic meaning is what the sign itself means. Now, the debate on DI is coined in the following terms: if a theory is able to derive linguistic meaning from speaker meaning, then DI would be vindicated. That is, if the a theory can reduce the fact that language has meaning to the fact that people usually intend to express certain things by using signs, then the intentionality of language would be derived from the intentionality of mind. This is the thesis we are going to analyze in more detail.

The most influential view implying DI has been the Gricean theory on meaning. According to a view based on Grice (1957, 1968, 1969):

(G)

(G1) A *speaker* S means that M by a sign E if S intends to produce a belief that M in an addressee by means of the recognition of the agent's intention.

(G2) A *linguistic sign* E means that M (E conventionally means M) if many members of the group mean M by uttering E

If this were true, the meaning of linguistic expressions would derive from the particular intentions that speakers have by using them (a version of DI). On this view on meaning, understanding a linguistic expression would be just to recognize

the speaker's intentions while using this sign. This is what I am going to call the thesis of Semantic Reduction (SR).

I think that not only this Gricean view (G) on language implies DI, but also that DI implies, at least, SR. The argument is the following: everyone accepts that for communication to be possible, the addressee must understand the meaning of the linguistic expression (P1). Now, suppose that DI is true and linguistic meaning can be reduced to speaker meaning (P2). It is also a shared assumption that speaker meaning is what speakers intend to express by a sign (P3). Therefore, by P1, P2 and P3 it follows that for communication to be possible, the addressee must come to know which are the speaker's intentions. Consequently, if DI is true understanding language is just understanding the speaker's intentions, that is, SR. This is important, because if we are able to show that G and SR are false, then DI can be reasonable rejected.[1]

This is the state of play. After stating the reductive approach, I am going to consider several difficulties with this account that directly concern our topic.

First of all, there is Searle's (1965) famous objection, which most people still accept as definitive, but which I think it is not a problem for G.[2] He suggests the case of an American soldier who gets lost in Italy during World War II. When he bumps into two Italian soldiers, he wants them to believe that he is a German officer, but the only words he remembers from German are 'Kennst du das Land wo die Zitronen blühen?' (Do you know the country where the lemon trees bloom?). The Italian soldiers do not know anything about German and come to believe that he is a German officer. As Searle points out, this is a serious problem for the Gricean theory as it was exposed in 'Meaning' (1957), because the account developed in this paper predicted that the linguistic meaning of the American soldier's utterance (what *is said*) would have been 'I am a German soldier', what seems blatantly wrong. However, according to G2, linguistic meaning derives from what people *usually* mean by it, (Grice (1968)) so Searle's objection does not work in our formulation of G. G does not predict that the American soldier's utterance means 'I am a German soldier', because that there is only one case in which someone means that by this utterance.

However, I do think that there are at least two sound arguments against G that strongly suggest that it cannot be right. First of all, this account is unable to satisfactorily account for literal meaning. Consider the classical sentence 'Could you pass me the salt'. Almost everyone uses this sentence intending it be a request (something like: *pass me the salt*) so, according to G2, its meaning is *pass me the salt*. However, we also want to say that it literally means something like 'Are you able to pass me the salt'. How can the G account for literal meaning? Notice that this is a compelling objection, for if linguistic meaning derives from speaker

[1] From the fact that G implies DI and G is false we can not directly conclude that DI is false, but if we add the premise that G as been the traditional view underpinning DI, the argument lends support to the rejection of DI.

[2] It must be said that G need not correspond exactly with the view of Grice (1957) (though I do think this is the theory resulting from Grice (1957, 1968, 1969)).

meaning, then the meaning of every sentence must be based on someone's intentions. If, in some cases, literal meaning is a kind of meaning that nobody ever intends to express, where does literal meaning come from?

It must be said that Grice foresaw this objection and replied that in some cases we know the meaning of a sentence if we know the procedure that *would* yield a meaning if someone intended to use it. However, this is just to give up DI, because at least in certain circumstances there is linguistic meaning without speaker meaning, and we still lack a good explanation for these cases. Furthermore, what prevents us from generalizing this account and explaining *all* linguistic meaning without relying on actual speaker's intentions?

Secondly, there is some empirical evidence that runs against SR (the thesis that understanding a linguistic expression is just recognizing the speaker's intentions), which is a thesis directly implied by G and DI. Consider what psychologists call a 'theory of mind', that is, the capacity to attribute mental states to oneself and to others and to interpret others behavior's in terms of mental states (Baron-Cohen, (1995)). Some experiments strongly suggest that autistic children (and some deaf children as well) lack a theory of mind. They never understand what it is for other people to have certain beliefs, intentions and desires and cannot explain other's behaviors in these terms. (Baron-Cohen (1995), Garfield *et al.* (2001)) Now, consider SR. If understanding language is just recognizing the speaker's intentions, people lacking a theory of mind would never master a language. However, the fact is that autistic children can utter and understand sentences and engage in a normal conversation. Arguably, they have problems with pragmatics (metaphors, irony,…), but still they have basic linguistic competence. Indeed, even non-autistic children master a language long before they attribute propositional attitudes to others, so that recognizing others' intentions cannot be a requirement for learning a language (Garfield *et al.* 2001). Consequently, G makes the wrong prediction.

Form these arguments I conclude that G cannot be the right account about the origin of linguistic meaning. The remainder of this paper is devoted to outline a framework which can explain the fact that language has meaning without relying on speaker meaning.

Notice that the project of explaining how meaning arises in language in this way depends on a more general project of naturalizing intentionality. The idea is that language can acquire its characteristic *aboutness* without deriving it from a more basic intentionality.

To see where the problem lies, we need to look back at G. G1 does not concern us; we can accept that it correctly defines what it is for someone to mean something by a sign. We are interested in G2, the claim that linguistic meaning is determined by speaker meaning. In other words, we want to understand what it is for language to *conventionally* mean something. We all agree that language is conventional. The problem with G may then lie in a wrong understanding of why conventions are and what it means for language to be conventional. So, to

understand how meaning can arise, we need first to understand what are we saying when we claim that language is conventional.

Millikan (1984, 2005) puts forward an account on conventions based on Lewis (1969) that tries to develop this idea. According to Millikan, for a convention to arise only two conditions need to be fulfilled: first, it has to be a *reproduction* of some past pattern. A reproduction is roughly defined as follows: x is a reproduction of y if and only if x is a copy o some aspects of y such that, had y been different, x would have been different accordingly. Secondly, the pattern of action that constitutes a convention must be *arbitrary*, that is, it does not have to be copied because it is the best solution to a problem, but just because of the weight of the precedents. To illustrate this idea, consider the convention of drinking green beer on St. Patrick's day. Drinking green beer is a pattern of action that gets reproduced just because other people use to drink green beer; if they had drunk red wine, people would act accordingly.

This basic account still needs to be complemented to explain the conventionality of language. According to Millikan,[3] some conventions have a further important feature: they help to achieve a common goal that both participants have. Consider shaking hands; when two persons meet, both are interested in introducing themselves; there are many ways to do it, (giving a hug, rubbing noses,…); what is important is not which action they choose, but that both participants choose the same one. Shaking hands is a reproduced and arbitrary pattern of action that keeps being copied because it achieves a common goal.

Language has all these features. Linguistic expressions are patterns of action that get reproduced (we copy the words and some structures), are arbitrary (there are thousands of languages) and are there because they carry out the function of communication. In particular, the common goal that language achieves is the production of true sentences and true beliefs. The hypothesis here is the following: if speakers stopped uttering true sentences and producing true beliefs addressees would stop listening; if, on the other hand, people stopped believing utterances, speakers would stop talking. In both cases language would die out (Millikan, 2005).

Now, how can this account help to explain linguistic meaning? The idea is that language is a convention, i. e. an arbitrary pattern of action that gets reproduced because it achieves communication. Now, the meaning of a linguistic expression is whatever there is in the world that explains why this particular expression gets reproduced. Take the case of 'water'; 'water' means water because this is what explains why the word 'water' keeps being reproduced in our languages. Of course, developing this idea lies far beyond the scope of this essay, but I hope the basic idea is clear: we need an externalist account of how language acquires

[3] This idea is also based on Lewi's account, but with an important difference. Lewis thought that all conventions must fulfill this condition, but Millikan criticizes (correctly, in my opinion) that some conventions do not achieve any goal. Consider whether there is a common goal in decorating the house at Christmas, or smoking a cigar when a new child is born.

meaning and, in my opinion, the Millikanian account on conventions is a promising way to do it.

Overall, my paper lends support to the rejection of DI, firstly by pointing at their problems and secondly by adopting a theory of conventions along the lines of Millikan, which constitutes an alternative account of the origin of meaning that implies the falsity of DI.

References

Baron-Cohen, S. (1995), *Mindblindness*, Cambridge, MA, MIT Press.

Garfield, J., C. Peterson and T. Perry, (2001), 'Social cognition, language acquisition and the development of the theory of mind', *Mind and Language* 16.5, pp. 494-541.

Grice, P. (1957), 'Meaning', *The Philosophical Review*, 66, pp. 377-88.

— (1968), 'Utterer's Meaning, Sentence Meaning, and Word-Meaning,' *Foundations of Language*, 4, pp. 225-42.

— (1969), 'Utterer's Meaning and Intentions,' *The Philosophical Review* 68, pp. 147-77.

Lewis, D. (1969), *Convention*. Cambridge, MA: Harvard University Press.

Millikan, R. (1984), *Language, Thought and Other Biological Categories,* Cambridge, MA, MIT Press.

Millikan, R. (2005), *Language: A Biological Model*. Oxford, Clarendon Press.

Searle, J. (1965). 'What is an Speech Act?' *Philosophy in America*, pp. 221-239.

— (1983), *Intentionality*, Cambridge, Cambridge University Press.

What does embodiment mean? Questioning the autonomy of psychology*

Saray Ayala López
Universitat Autònoma de Barcelona
sarayayala@gmail.com

Introduction

The Multiple Realizability thesis (MRT) (Putnam, 1967) has been taken to warrant the autonomy of high-level sciences such as psychology. Fodor (1974) argued that since properties in higher-level sciences are multiply realizable in properties in lower-level sciences, generalizations in higher-level sciences have no physical counterparts, and therefore higher-level sciences are autonomous.

Here I will defend an argument against the autonomy conclusion that draws into question the idea that psychological generalizations are alien to details of embodiment. Under the embodied cognition framework, cognition depends on body and environment. I will argue for one interpretation of this dependency according to which a psychological description of an organism's mind cannot be made in the allegedly autonomous way that psychology is supposed to provide.

Embodied cognition

Senses of embodiment

Embodied cognition is a broad research program developed in different fields with a common emphasis against traditional cognitive science (Anderson, 2003; Calvo and Gomila, 2008). This trend places a new emphasis on the active role body and environment play. For our purposes, we will refer to embodied cognition (EC) as claiming that (i) cognition depends on the body, and that (ii) cognition depends on the environment. For the sake of simplicity, I will focus on (i).

We can identify two main possible contributions from body to cognition. First, a *causal story* claims that (some) bodily processes can cause (some) cognitive processes (See Aizawa, 2007). Second, a *dependency story* proposes a stronger contribution of bodily events in cognitive processes to be cashed out in terms of dependency relations (See Noë, 2004). This dependency contribution can be understood in two different ways, depending on whether we focus upon *implementation* or *computation*. What I label the *implementational story* claims that organism meet the *principle of total embodiment* (PTE). A system meets PTE

* Preparation of this work was made during my (delightful) research stay at the University of British Columbia, supported by a La Caixa Scholarship. I would like to thank Paco Calvo for his valuable, encouraging and fun comments.

only if some cognitive processes depend partly on specific bodily details of implementation of the system. What can be called the *computational role story* proposes that the body plays a computational role in cognition, although its (physical) details are not important. By *computational role* we refer to a role as described at David Marr's *representational and algorithm level* (Marr, 1982).

Research on EC may tell against the autonomy of psychology in the following way:

(1) If EC is correct, cognition depends upon body and environment

(2) If cognition depends upon body and environment, psychology is not autonomous

(3) EC is correct

(4) Cognition depends upon body and environment

Therefore,

(5) Psychology is not autonomous

Morphological computation

Research in EC illustrates how the body can shape the mind of a system in the strong sense that is our concern here[1]. To exemplify the idea of embodiment I present a robot (Paul, 2004), that in spite of being controlled by two perceptrons, unable to compute linearly inseparable functions as exclusive-OR (XOR), gets to exhibit some form of XOR-constrained behavior. And this is courtesy of its morphology[2].

The inputs coming into the two perceptrons are A and B. One network computes OR and is connected to a motor M1, the other computes AND and is connected to a motor M2. M1 turns a single wheel causing forward motion, whereas M2 serves to lift the wheel off the ground. The interesting case is when A and B are both active. The OR network makes M1 to move but the AND network lifts the wheel from the ground, so the robot remains stationary (Fig. 1). Summarizing the behavior of the robot in a table, we discover that it looks like the truth table of the XOR function (Fig. 2).

We can say that Paul's robot is a case of the *dependency story*. The specific sense of this dependency will be considered in the following sections[3].

Two stories

[1] See Spivey *et al*. 2004; Ballard *et al.* 1997; Noë, 2004; O'Regan & Noë, 2001; Lakoff & Johnson, 1999; Hurley, 1998

[2] To be a linearly separable or inseparable function refers to the possibility of drawing a line, in a spatial representation of that function, which divides the representational space according to the physical similarity of the input patterns.

[3] Someone might complaint about the simplicity of our example. Paul's robot is relevant here because it solves a function that requires a level of processing that abstract away from physical similarities among inputs. And this level of abstraction is what allows us to talk of semantics and cognition when human perform these kinds of abilities.

Extended Functionalism vs. Body-determinism

Extended functionalism (Clark, 2006, 2007) is an extension of (classical) functionalism where mental states are functional states (Fodor, 1975; Putnam 1975). The important criterion for mental sameness is not physical-sameness, but functional-sameness. Body-determinism, however, claims that details of the body do matter (Shapiro 2004, 2006).

The body-determinist would say that the robot meets PTE. According to the functionalist, however, the morphology of Paul's robot plays a particular computational role, its specific physical details being unimportant. The same computational role can be played by several different elements. Let's consider the standard (disembodied) computation of XOR. Functions OR and AND can be computed with two input units and one output unit (Fig. 3 and 4). Computing XOR requires another (hidden) unit (Fig. 5). This hidden unit becomes active when both inputs are active, sending a negative activation to the output unit equivalent to the positive activation that it receives from the input units. Let this standard approximation of XOR be case 1, and let call Paul's robot case 2. According to a functionalist, case 1 is equivalent to case 2 in the relevant sense: in both cases the description at the computational level is the same. The difference can be displayed as a difference in implementation. The robot's morphology can be used as a computational unit, as if it was the hidden unit in a standard three-layer network. According to body-determinism, however, cases 1 and 2 are not computationally equivalent[4]. Computing XOR is not the same as computing AND and OR with a particular morphology.

Against Body-determinism

Friends of the (extended) functionalist position can argue against body-determinism at least in two ways: one empirical and one conceptual. The empirical strategy would be to claim that the evidence provided by body-determinism is not enough to support its claim that organisms meet PTE. The conceptual strategy would consist of exploiting a more general anti-reductionist strategy. Since we can imagine cognition being realized in other devices different from the human brain and body, then cognition is not tied to specific meat. That is, they appeal to MRT.

Replies

Against the empirical strategy

What are the possibilities for body-determinism in order to respond to the empirical strategy? Is it possible to show that an organism meets PTE? Under the functionalist lens, the response is negative: any bodily contribution can be seen as performing a computational role, the (physical) details being unimportant for the achievement of this role. This supposed empirical strategy is actually non-

[4] *Computationally equivalent* means here a strong equivalence. Not only exhibiting the same behavior given a particular set of inputs, but equivalence in the computational steps.

empirical. Even more, it begs the question, because it assumes that mental states are multiple realizable, that is, that organisms do not meet PTE.

Against the conceptual strategy

Following the scheme made by Marr (1982), functionalists elaborate the functional description of a cognitive process with no mention of the physical details or the computational steps performing it. At the algorithmic level they specify the computational operations that realize it, assuming that there are representational states inside the system that can stand for a variety of different states. The functionalist cognitive function is free from any physical details, it is said to (be able to) be the same whatever the details of implementation are. The reductionist cognitive function is, however, constrained to a particular type of body. But extended functionalism also seems to be constrained by something: it assumes a disembodied computational description of mind. Let me elaborate this.

According to the functionalist, Paul's robot case only shows that morphology can play a computational role, and it does not demonstrate that its particular morphology is critical[5]. In fact we can easily think of various robots with different morphologies that achieve a XOR-constrained behaviour. What we may call XOR-robot-2 is like Paul's except for the fact that there is no lift. Instead, M2 just liberates the wheel from the propulsion from M1, leaving the wheel now free, although passively driven in case of any other motion. These robots, different in morphology, are however equivalent at the behavioral level. Let's consider again the standard computation of the XOR function. The hidden unit, functionalist says, is playing the same computational role that the lift plays in Paul's robot and the brake in the XOR-robot-2 (i.e. some sort of constrain when both inputs are active). Robots' bodies are then (just) playing the appropriate computational role.

Functionalists endorse, as we see, the same (disembodied) computational description (i.e. the XOR function as described by its truth table) across different ways of realizing it. And that is possible only if you hold a disembodied notion of cognition and assume that inside the system there are representational states that can be realized by different states, inside or outside the skull (i.e. the hidden-unit in a standard XOR network, the lift in Paul's robot, etc). But if you hold an embodied view, then you cannot describe the cognitive function independently of the particular computations that are happening to occur (that is, independently of whether the behaviour is achieved by means of computing XOR with a three-layer feedforward network, or by means of computing AND and OR with a particular morphology).

Paul's robot exhibits a XOR-constrained behaviour while the floor is even. In case there is a level change so that the lifted wheel gets to the ground, when A and B are active the robot will move forward, and then it will fail to achieve XOR. If XOR-robot-2 is placed downhill the function will also change as a whole: the

[5] According to the extended functionalist, Paul's robot is considered as evidence in favor of the extended functionalist reading.

slope will naturally drive the liberated wheel and the robot will move. With these examples we see that the functions that describe robots' behaviour are affected as a whole by the environment, because of their particular morphology. In explaining why the robots fail to get the XOR–constrained behaviour in these cases, we realize that we cannot describe the function independently of the way in which that function is actually being realized. How would a sympathizer of functionalism explain these cases? When everything is OK and our robots get the XOR-constrained behaviour, it seems that they are sharing a common way of achieving it, because the description at Marr's computational level is the same. The only difference among them seems to be at the implementational level. But when something is wrong (e.g. a floor change) then we see that robots fail to accomplish the cognitive task we described at Marr's computational level because at the algorithmic level nothing is being shared.

The sympathizer of extended functionalism is assuming a particular disembodied computational description of the mind across all possible bodies and environments. But which is that function that can be realized with different computational steps and in many different implementations? If I am right, there is no such independent-of-the-details function.

If these considerations are on the right track, the functionalist is obliged to choose between the following options:

1. If she wants the benefits of a multiply realizable disembodied description of the mind, then she has to give up the embodied framework and go back to the traditional view

2. If she wants the benefits of the embodied mind, then she has to give up the functionalist fantasy of maintaining one and the same cognitive function at the computational level independently of the particular details of the body and the environment

From body-determinism to the non-autonomy of psychology

According to the considerations made above, body-determinism seems to be the best way to understand EC. We have been focusing on the body, but, as we said at the beginning, EC also claims that cognition depends on environment. Our conclusion about bodily details is applicable to environmental details. The debate is the same: do physical details matter? We can expand body-determinism by claiming that physical details of embedment matter as well. Our argument against the autonomy of psychology goes as follows:

> (i) EC claims that mind depends upon bodily and environmental details
>
> (ii) Body-determinism seems to be the best way to understand this dependency
>
> (iii) If body-determinism is right, an explanation of an organism's mind cannot be made without describing bodily and environmental implementational details

(iv) Psychology is said to be autonomous because it can describe an organism's mind independently of that organism's implementational details

(v) It follows from (iii) that a psychological explanation of an organism's mind needs to include bodily and environmental implementational details

Therefore,

(vi) Psychology is not autonomous

References

Aizawa, K. (2007), 'Understanding the Embodiment of Perception', *The Journal of Philosophy* Vol. CIV, 1, pp. 5-25.

Anderson, M. (2003), 'Embodied Cognition: A field guide', *Artificial Intelligence* 10, p. 10

Calvo, P. and Symons, J. (2008), 'Radical Embodiment and Morphological Computation: Against the Autonomy of (some) special sciences, in *(2008) Reduction and the Special Sciences (Tilburg, April 10-12, 2008)*.

Calvo, P. and Gomila, T. (eds.) (2008), *Handbook of Embodied Cognitive Science*, Elsevier.

Clark, A. (2006), 'Pressing the flesh: Exploring a tension in the study of the embodied, embedded mind', *Philosophy and Phenomenological Research* 76, 1, pp. 37-59.

— (2007) 'Curing the Cognitive Hiccup: A Defense of the Extended Mind', *Journal of Philosophy* 104, p. 4.

Fodor, J. (1974), 'Special sciences and the disunity of science as a working hypothesis', *Synthese* 28, pp. 77-115.

Lakoff, G. and Johnson, M. (1999), *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*, New York, Basic Books.

Marr, D. (1982), *Vision,* San Francisco, W. H. Freeman.

Noë, A. (2004), *Action in Perception*, Cambridge, MIT Press

O'Regan, J. and Noë A. (2001), 'A sensory motor approach to vision and visual consciousness', *Behavioral and Brain Sciences* 24, pp. 939-73.

Paul, C. (2004), 'Morphology and Computation', *Proceedings of the International Conference on the Simulation of Adaptive Behaviour*, Los Angeles, CA, USA, pp. 33-38.

Putnam, H. (1967), 'Psychological Predicates', reprinted in Block (1980) and elsewhere as 'The Nature of Mental States.'

— (1975), 'Philosophy and Our Mental Life', in H. Putnam (ed) *Mind, Language and Reality*, Cambridge University Press, Cambridge, UK, pp. 291-303.

Shapiro, L. (2004), *The Mind Incarnate*, Cambridge, MIT Press.

— (2006), 'Reductionism, Embodiment, and the Generality of Psychology', in H. Looren de Jong and M. Schouten (eds.), *The Matter of Mind*, Malden, MA, Blackwell Publishing, pp. 101-20.

Figures

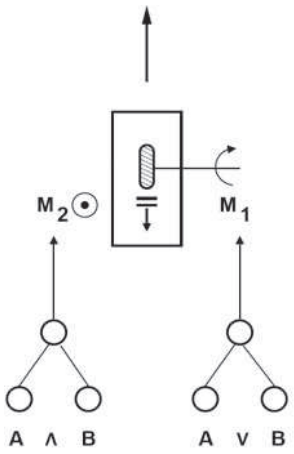


Figure 1: XOR- Robot. From Paul, C. (2004), p. 2

A	*B*	Behavior
F	F	stationary
F	T	moving
T	F	moving
T	T	stationary

Figure 2: XOR-Robot behaviour. From Paul, C. (2004)

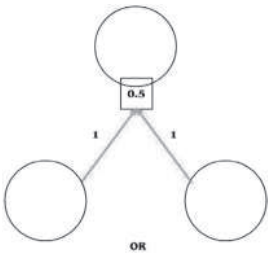


Figure 3: OR network

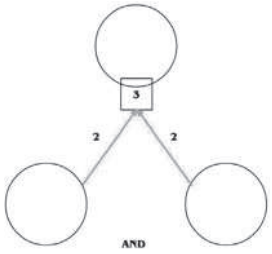


Figure 4: AND network

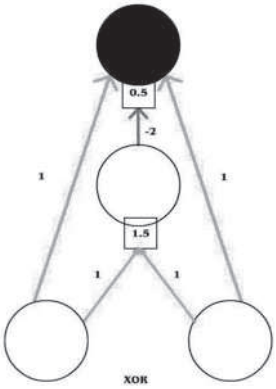


Figure 5: XOR network

Eliminativismo perlocucionario*

Antonio Blanco Salgueiro
Universidad Complutense de Madrid
ablancos@filos.ucm.es

Defenderé que la noción de *perlocución* y la distinción *ilocucionario/perlocucionario*[1] son confusas, por lo que deben ser *eliminadas* de la Teoría de Actos de Habla (TAH) y *sustituidas* por conceptos teóricos mejor perfilados. La caracterización inicial de las perlocuciones en Austin (1962) es:

> "There is a further sense (C) in which to perform a locutionary act, and therein an illocutionary act, may also be to perform an act of another kind. Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons: and it may be done with the design, intention or purpose of producing them." (p. 101)

La literatura posterior revela tres tipos de problemas: respecto a los "efectos perlocucionarios", las "causas perlocucionarias" y la "causación perlocucionaria".

Efectos perlocucionarios

¿Se reducen a efectos sobre el oyente, o incluyen efectos sobre el hablante o terceras personas, y efectos sobre el mundo físico y social? Aunque Austin admite otras posibilidades, sus ejemplos son de efectos sobre la audiencia. Pero el habla produce (intencionadamente o no) efectos sobre el propio hablante que pueden reformularse en el lenguaje de la acción: "Por insultarlo, me desahogué"; "Por confesar, me alivié". En cuanto a efectos sobre el mundo, al decir que Bush atacó Irak hablamos de una acción que consideramos efecto de una ilocución (su *orden* de atacar).

En lo que se refiere al tipo de efectos involucrados, Austin distingue entre sentimientos, pensamientos y acciones, pero pueden incluirse otros estados mentales (deseos, intenciones, emociones, etc.), lo cual sugiere una taxonomía basada en el tipo de efecto mental producido (Gaines, 1979).

Las cosas se complican cuando Austin afirma que algunos efectos son constitutivos del acto ilocucionario, y señala tres tipos de "efectos ilocucionarios":

> 1) La comprensión del oyente. Para Austin ninguna ilocución consiste en un acto unilateral del hablante. El oyente debe contribuir al menos con su

* Este trabajo ha sido financiado por los proyectos HUM2006-04955/FISO (Ministerio de Educación y Ciencia) y FFI2008-03902 (Ministerio de Ciencia e Innovación).

[1] Austin (1962), Cohen (1973), Campbell (1973), Gaines (1979), Davis (1979), Nicoloff (1989), Gu (1993), Kurzon (1998) y Cavell (2005).

comprensión. Si *H* dice "Vendré mañana" y *A* no lo toma como una promesa, *H* no ha prometido.

2) "Tener efecto". Un bautizo hace que ciertas prácticas de denominación sean legítimas y otras no. Nombrar a alguien para un cargo unipersonal hace que esa persona lo ocupe y que ninguna otra pueda ocuparlo.

3) Muchas ilocuciones invitan (convencionalmente) a una respuesta o secuela. ¿Cómo caracterizar una *protesta* sin aludir a que se busca una rectificación, o una *orden* sin aludir a actos de obediencia perseguidos, etc.? Este punto pone claramente en entredicho la pretensión de trazar una frontera nítida entre lo ilocucionario y lo perlocucionario. Cohen (1973) distingue tres clases de efectos perlocucionarios: i) Efectos producidos por los aspectos locucionarios, sin mediación de los aspectos ilocucionarios, como cuando digo "¡No te despiertes!" causando que alguien se despierte (causa fonética), o llamo la atención de alguien porque pronuncio su nombre (causa rética). 2) Efectos producidos por mediación de la ilocución (causa ilocucionaria), pero no "asociados" a ella, como *sorprender*, *asustar*, *divertir, aburrir, exasperar* o *fascinar*, en relación a *afirmar*. Al *afirmar* algo puedo *sorprenderte*, pero ello no tiene que ver con lo que esencialmente es afirmar algo. 3) Efectos "asociados" a la ilocución (causa ilocucionaria), como intimidar con respecto a amenazar, alertar con respecto a advertir, persuadir o disuadir con respecto a argüir, obedecer con respecto a ordenar, o responder con respecto a preguntar. Estos efectos son los que Austin tenía en cuenta al señalar el tercer sentido en el que las ilocuciones están vinculadas con la producción de efectos. El objetivo de producirlos debe mencionarse para caracterizar la correspondiente ilocución, ya que constituyen su *objetivo* (*point*). Aquí se aplica la *teoría de los infortunios*. Hay algo desafortunado en una pretendida protesta que no quiera cambiar nada, en una orden que no deba ser obedecida, en una amenaza que no busque intimidar. Lo constitutivo es el objetivo de producir el efecto, más que su producción efectiva, pero lo importante es que no se puede separar el análisis de ciertas ilocuciones del de ciertas perlocuciones.

"Causas" perlocucionarias

Cabe discutir si son siempre ilocuciones o cuentan también las locuciones (en sus aspectos fonéticos, fáticos o réticos). Cohen denomina "oblicuas" a las perlocuciones causadas por locuciones, y "directas" a las causadas por ilocuciones.

Tipo de causación

Cabe discutir si es físico-mecánica, verbal (mediada a través del lenguaje y la comprensión), o verbal-influencial (mediada además por motivos o razones), o incluso si es legítimo hablar aquí de causación (Gu, 1993). El pasaje clave lo encontramos en la p. 113 de Austin (1962):

> "the sense in which saying something produces effects on other persons, or *causes* things, is a fundamentally different sense of cause from that used in

> physical causation by presure, &c. It has to operate through the conventions of language and is a matter of influence exerted by one person on another"

Esto permitiría eliminar algunas de las perlocuciones "oblicuas", dejando sólo a las "directas", pero no permite distinguir, entre éstas, las asociadas de las no asociadas.

También se discute si la causación debe ser intencional o puede no serlo, como admite Austin. Autores, como Gaines (1979) o Bach & Harnish (1979), se muestran partidarios de un *intencionalismo perlocucionario*, porque para ellos la pragmática debe ocuparse sólo de los objetivos comunicativos del emisor.

Las ambigüedades señaladas apuntan a una gran heterogeneidad dentro de lo "perlocucionario". La reivindicación de algunos de los fenómenos enredados en esa madeja podría venir de marcos teóricos que justificasen la necesidad de prestarles atención[2].

Al final de la conferencia VII Austin (1962) abandona la distinción constatativo/realizativo y propone una *realizatividad generalizada.* Si antes sostenía que *a veces* hablar es actuar (pero otras es constatar), ahora asume que hablar *siempre* es actuar, y que la acción que típicamente realizamos al hablar es compleja, estando estructurada en varios actos y sub-actos entrelazados.

> **RG: Realizatividad generalizada:** *Hablar es realizar actos.*

Por otra parte, asume una máxima holística (Austin, 1962, p. 148):

> **H: Holismo:** *El acto de habla total en la situación de habla total es el único fenómeno real que en última instancia estamos tratando de aclarar.*

Mi crítica a la noción de perlocución proviene de prestar atención a una tensión entre **RG** y **H**. Mientras **RG** dicta que el habla puede explicarse exhaustivamente como una estructura de actos entreverados, **H** incorpora el importante concepto de *situación de habla total*, como distinta del *acto de habla total.* Austin acaba primando **RG**, lo cual le hace perder de vista la importancia de la *situación* de habla. El "mito de lo perlocucionario" hace que resalten los aspectos del habla que se dejan tratar, aunque sea forzadamente, como actos, y vuelve invisibles para la mirada teórica otros aspectos que poseen una importancia equivalente pero no se dejan tratar como tales. Presentaré algunos ejemplos ilustrativos, y una alternativa en la que es central la noción de *marco*. Llamo *marco asociado* a los antecedentes y consecuencias cuya mención es necesaria para explicar la propia ilocución, mientras que el *marco no-asociado* contiene antecedentes y consecuencias que no es preciso mencionar para caracterizar la ilocución, aunque pueden ser parte de la situación de habla total en casos particulares. La distinción entre componentes del marco "asociados" y "no asociados" no es nítida, ya que algunos componentes que no están estrictamente asociados a la ilocución pueden formar parte de situaciones de habla prototípicas. Los dos primeros ejemplos muestran sólo elementos del marco asociado. El tercero incluye también algunos elementos de un marco no asociado (típico) (*H*: hablante; *A*: Audiencia).

[2] En Campbell (1973) se liga el estudio de las perlocuciones a la retórica (véanse también Gaines, 1979 y Gu, 1993).

antecedentes		**ilocución**		**consecuencias**
H se siente *ofendido* por *A*	→	*H insulta* a *A*	→	*A* se siente *ofendido* por *H*
H se siente *disgustado* por la acción *a* de *A*	→	*H protesta*	→	*A rectifica* *A se disculpa*
H: generosidad; sentirse en deuda con *A*; aprecio por *A*; creer que *A* desea *p*				obligación de hacer *p*
		→ *H promete* que *p* →		
A: deseo de que *H* haga *a,* creer que *H* le debe algo, sentir el aprecio de *H*				alegría; expectativa de que *H* haga (o intente hacer) *p*; agradecimiento;

Una TAH que adopte la noción de perlocución hace aparecer, en todos esos casos, las consecuencias (reformuladas como actos) como parte de su estudio, mientras que considera los antecedentes que no se dejen tratar como condiciones ilocucionarias como fuera del dominio de la pragmática. Pero ¿por qué la pragmática va a estar menos interesada en la emoción que induce a *H* a insultar a *A* que en la emoción que resulta en *A* como efecto del insulto? La decisión parece arbitraria: o ambas deben ignorarse (limitando el alcance de **H**), o ambas deben incluirse. Si se incluyen ambas, debe hacerse de modo que no las coloque en categorías completamente diferentes, ya que lo que parece involucrado en ambos casos es que el habla está atravesada por emociones (Cavell, 2005) y otros estados mentales. Algunos de los antecedentes son tratados en la TAH "ortodoxa" como parte del estudio de las ilocuciones, como condiciones de sinceridad o condiciones preparatorias. Esa es una razón adicional para introducir la noción de marco. La TAH hace aparecer en lugares dispares (condiciones ilocucionarias/efectos perlocucionarios) elementos de naturaleza similar. Mi propuesta consiste en abandonar la noción de perlocución y sustituirla por la de *marco de habla*. En cuanto a si tanto los marcos asociados como los no asociados deben formar parte de un estudio pragmático del lenguaje, ello depende de cómo se conciba dicho estudio (y de la solidez de la distinción asociado/no asociado). Tienen especial interés los marcos que son imprescindibles para explicar las ilocuciones, los marcos asociados. Propongo denominarlos *marcos ilocucionarios*, ya que su estudio formaría parte de la teoría de la fuerza. El primer ejemplo recoge la idea de que los insultos son ilocuciones motivadas por un estado de animadversión hacia el insultado y que, a la vez, intentan inducir un estado mental negativo en éste. No podríamos entender el *uso insultivo* del lenguaje sin hacer referencia a un marco como ése.

Una razón adicional para no tratar la parte derecha de un marco como *actos* del hablante es que ello produce una visión monologista de la situación total del habla

en la que la audiencia sólo se introduce de contrabando. Tratar la contribución de la audiencia como algo importante de suyo requiere tratarla como parte de la situación total de habla, no como una parte supeditada a la acción del hablante[3]. Considerar la respuesta de la audiencia como un mero efecto de la emisión del hablante la hace aparecer como pasiva, cuando lo que ocurre en casos paradigmáticos (como *convencer*) es que la audiencia es la principal responsable del evento en cuestión.

Referencias bibliográficas

Austin, J. L. (1962), *How to do Things with Words*, Oxford, Oxford University Press.

Bach, K. y Harnish, R. M. (1979), *Linguistic Communication and Speech Acts,* Cambridge, MA, MIT Press.

Campbell, P. N. (1973), 'A rethorical view of locutionary, illocutionary and perlocucionary acts', *Quarterly Journal of Speech* 59, pp. 284-96.

Cavell, S. (2005), 'Performative and passionate utterance', en S. Cavell, *Philosophy the Day After Tomorrow,* Cambridge/Londres, Harvard University Press, pp. 155-91.

Cohen, T. (1973), 'Illocutions and perlocutions', *Foundations of Language* 9, pp. 492-503.

Davis, S. (1979), 'Perlocutions', en J. R. Searle, F. Kiefer y M. Bierwish (eds.), *Speech Act Theory and Pragmatics,* Amsterdam, Reidel, pp. 37-55.

Gaines, R. N. (1979), 'Doing by saying: toward a theory of perlocution', *Quarterly Journal of Speech* 65, pp. 207-17.

Gu, Y. (1993), 'The impasse of perlocution', *Journal of Pragmatics* 20, pp. 405-32.

Kurzon, D. (1998), 'The speech act status of incitement: perlocutionary acts revisited', *Journal of Pragmatics* 29, pp. 571-96.

Nicoloff, F. (1989), 'Threats and illocutions', *Journal of Pragmatics* 13, pp. 501-22.

[3] El "monologismo" perlocucionario de Austin es señalado por Campbell (1973) y Gu (1993).

Elusive and holistic transfer of warrant: The externalist new *cogito*

Cristina Borgoni Gonçalves and Manuel de Pinedo García
Universidad de Granada
cborgoni@yahoo.com.br / pinedo@ugr.es

Moore's proof of the external world has received much attention during the last few years within the context of the debate regarding transfer of warrant. We will start by applying some ideas from David Lewis (1996) to the diagnoses offered by Wright [(2000), (2003)] and Davies [(2000), (2004)] questioning the success on the transmission of warrant in the proof. Then we will wonder whether a similar approach is apt for variations of the proof, such as the argument against the compatibility between externalism and self-knowledge [McKinsey (1991), Boghossian (1998)]. We will argue that Lewis' ideas need to be complemented by a rejection of some atomistic commitments regarding mental content common to some externalists and incompatibilists. The alleged incompatibilist argument becomes a compatibilist one (and, to our minds, correct) once externalism is accompanied by holism regarding mental content. We will conclude that "I think, therefore there is an external world".

Here is Moore's proof [Moore (1939)]:

(M1) I have two hands: here is one hand, here's another,

(M2) If I have two hands, then the external world exists,

Hence: (M3) The external world exists.

A very popular reaction to this proof is to question our entitlement to affirm M1 if we are not previously justified in affirming M3, either because M3 is part of the epistemic warrant for M1, or because M1 and M3 are in need of previous justification [see Wright (2000), (2003) and Davies (2000), (2004)]. This approach to the proof can be expanded in light of Lewis' defence of contextualism (1996). According to Lewis, when we move from M1 to M2 we are performing a context shift: while we are fully justified in stating M1 in standard situations, if the situation envisaged is one where questioning the existence of the external world makes sense (and an academic discussion on epistemology is such a situation), then such a possibility can no longer be ignored and we cannot say anymore that we know we have two hands. Lewis' take is, in this sense, a variation on the reaction above: M1, said in an epistemological context, is not warranted unless we have means to reject the negation of M3.

Let's consider now whether the challenge against the compatibility between externalism and self-knowledge could be tackled in a similar vein. The claim is that the *a priori* character of the following argument is a *reductio* of compatibilism (Boghossian 1998):

(I1) I think that water is wet,

(I2) If I think that water is wet, then there is water in my environment,

Hence: (I3) There is water in my environment.

Is there a context shift between I1 and I2? One could argue that the externalism that is made explicit by the conditional I2 means that the warrant for I1 does not transfer to I3: I1 could be justified even for an internalist, but the introduction of I2 means that we would only be justified in stating I1 (now forced to give it an externalist reading) if we were already justified to state I3. However, it is not clear at all that there is a shift of context between I1 and I2 parallel to that between M1 and M2; for once, the possibility of scepticism doesn't seem to be involved in asserting or questioning I2 [see Sawyer (2006)]. Externalism is not primarily directed to refute scepticism. Under Lewis' approach, doubting (I3) is reasonable in an epistemological context in contrast to our ordinary life. But this doesn't mean that such a doubt should be taken into account in every philosophical context (for instance, in one concerned with the nature of our minds).

Nevertheless, we believe that there is a deeper problem with the present argument. It is not externalism *tout court* that justifies I2, but a very specific—although very influential—form of externalism. One could assert I2 only if committed to the idea that the relevance of the environment for the individuation of concepts is merely piecemeal (concepts making reference to the environment are individuated by the features of the environment they are about).

But, of course, such an atomistic take on externalism is optional. Besides highlighting the role of certain aspects of the environment for certain concepts, a more holistic form of externalism would claim that no area of a person's mental life can act as an independent variable with respect to the external world and, in parallel, that even concepts that fail to establish a connection with the world (say, phlogiston or god) have external identity conditions inasmuch as they belong to a conceptual network which is unintelligible independently of the world [for a defence of this kind of externalism, see Borgoni (2009)]. A non-atomistic externalism would not warrant the assertion of I2 and would not be prone to the accusation of incompatibility with self-knowledge.

And yet, the incompatibilist could mount an apparently more devastating *reductio* argument against this form of externalism. The argument could look like this:

(I'1) I think (for instance, that water is wet, or that the external world doesn't exist),

(I'2) If I think, the external world exists,

Hence: (I'3) The external world exists.

If the deduction of the existence of water from the mere fact that I have thoughts involving the concept 'water' is perplexing, the deduction of the existence of the external world from the mere fact that I have any given thought should be even more so. Or should it? We are ready to bite the bullet and to claim that a proper understanding of thought, holistic and fully externalistic implies, from the start,

the existence of reality. To put it in slightly different terms: the very intelligibility of thought (including false thoughts, or thoughts containing empty concepts) depends on the existence of the world. *Cogito ergo mundus est*.

References

Boghossian, P. (1998), 'What the Externalist Can Know 'A Priori'', *Philosophical Issues* 9, pp. 197-211.

Borgoni, C. (2009), 'En casa, en el mundo: el externismo global constitutivo', *Teorema* XXIII/3, en prensa.

Davies, M. (2000), 'Externalism and Armchair Knowledge', in Boghossian, P. & Peacocke, C. (eds.), *New Essays on the A Priori,* Oxford, Oxford University Press, pp. 384-414.

Davies, M. (2004), 'Epistemic Entitlement, Warrant Transmission and Easy Knowledge', *Proceedings of the Aristotelian Society*, Supplementary Volume 78, pp. 213-45.

Lewis, D. (1996), 'Elusive Knowledge', *Australasian Journal of Philosophy* 74, pp. 549-67.

Mckinsey, M. (1991), 'Anti-Individualism and Privileged Access', *Analysis* 51, pp. 9-16.

Moore, G. E. (1939), 'Proof of an External World', *Proceedings of the British Academy* 25, pp. 273-300.

Sawyer, S. (2006), 'Externalism, Apriority and Transmission of Warrant' in Marvan, T. (ed.), *What Determines Content? – The Internalism/Externalism Dispute*, Cambridge, Cambridge Scholars Press, pp. 142-153.

Wright, C. (2000), 'Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof'', *Philosophical Issues* 10, pp. 140-63.

Wright, C. (2003), 'Some Reflections on the Acquisition of Warrant by Inference' in Nuccetelli, S. (ed.), *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge Mass, MIT Press, pp. 57-77.

The temporal grounding problem in light of different notions of object

Marta Campdelacreu Arqués
Universitat de Barcelona & LOGOS
marta.campdelacreu@gmail.com

Introduction

Let us focus our attention on the philosophically famous statue Goliath made out of the piece of clay Piece. Let us suppose that they are created simultaneously at t_1, they coincide for a while and, at t_2, we squash them and, therefore, Goliath gets out of existence. Piece is still there, but now it is not coincident with Goliath.

Two prominent theories about the relationship between Goliath and Piece (and so about the relationship between ordinary middle-sized material objects and the pieces of matter out of which they are made) are *endurantism* and *perdurantism*. *Grosso modo*[1], perdurantism is the thesis that objects persist through time by having a temporal part at every time they exist; endurantism is the thesis that objects persist through time by being *wholly present* whenever they exist.

It has been argued that the phenomenon of temporal coincidence exemplified above gives rise to the so-called '*temporal grounding problem*' which would be easily explained or dissolved by perdurantist theories, but that would be a real difficulty for endurantist theories.

This is the *temporal grounding problem*: Goliath (the statue) and Piece (the piece of clay) are two different objects[2] that, as we have seen, are coincident at a given time and therefore share a lot of their properties at that time: their material components (at least at some basic level of composition), their mass, shape, colour, texture, etc. The problem is that, as Katherine Hawley says, in [Hawley (2008)], these properties are thought to be *basic* in the sense that they seem to determine the properties that, alas, Goliath and Piece do not actually share, such as their sortal, modal, aesthetic or futural properties. How may this situation be explained? (I will focus on sortal properties.)

In this paper I will argue that both endurantism and perdurantism can solve the temporal grounding problem equally well but that they cannot do it in a parallel way (as has sometimes been argued), as the notions of object that the two theories presuppose are radically different. In so doing, I will highlight some lessons that

[1] Later I will need to be more precise about the versions of these theories I am considering.
[2] This makes clear that the temporal grounding problem affects those versions of perdurantism which accept that, for example, Goliath and Piece are two different objects. The *Stage View*, for example, is not affected by it. In this paper I will only consider the version of perdurantism that accepts that Goliath and Piece are two different objects.

endurantists can learn from the temporal grounding problem about the notion of object that they are "building".

A perdurantist solution to the temporal grounding problem

Perdurantists (of the kind under consideration here) deny the intuition behind the temporal grounding problem and maintain that at the time of coincidence we have two objects, Goliath and Piece, which share quite a lot of their properties but differ in their sortal properties, among others. They explain the situation by appealing to several features of their notion of object.

First, they appeal to the idea that objects are just partially present at the different times at which they exist. An object partially present at a given time and place may have, at that region, its sortal property at least partly determined by other properties of the object that are temporally extrinsic to that region, and perhaps by other properties of the object's environment (intentional properties of people, properties having to do with the way in which the object has been created or is related to other objects...).

When we apply this idea to our example, we see that the temporal parts present at the time of coincidence are piece-of-clay-wise related but not statue-wise related to other temporal parts outside the time of coincidence. This would explain why Goliath and Piece are two different objects (they have different temporal parts) with two different sortal properties.

In addition to this, perdurantists also appeal to a second feature of their notion of object: that objects have temporal parts whenever they exist, and that these temporal parts can be shared by other objects. This second idea explains why even if objects like Goliath and Piece are different, they share so many properties at the time of coincidence: they share their temporal parts during that period.

Some lessons for endurantists

Now, what can we ask of an endurantist answer to the same problem? As well as perdurantists, endurantists deny the intuition behind the temporal grounding problem and maintain that at the time of coincidence we have two objects, Goliath and Piece, which share quite a lot of properties but differ in their sortal properties, among others. This being so, it seems reasonable to ask of endurantism the same kind of answer that perdurantism offers. This consists, I think, in an explanation of the resultant situation in terms of the endurantist notion of object. And, against what is sometimes claimed, this does not imply that the endurantist answer has to be given in terms parallel to those used to formulate the perdurantist answer. In particular, this does not imply that endurantists have to offer an explanation according to which the fact that the two objects have different properties outside the time of coincidence plays a crucial role. This is indeed the kind of explanation offered by perdurantists. However, it would be completely unfair to impose on alternative theories, with (may be radically) different views on what it means to be

an object, one such strategy —as a necessary condition for an answer to the temporal grounding problem to be correct.

This being said, I think that endurantists can perfectly explain the situation using some features of their notion of object. Let me explain.

One of the ideas often mentioned when characterizing endurantism is that objects are *wholly present* whenever they exist. This would mean, at least partly, that at every time of its existence, central features of an object (like its sortal property) are determined without appealing to futural properties of the object itself (however, they may be (partly) determined by properties of the object's environment). In our case, this would mean that, at the time of coincidence, Goliath/Piece have their sortal properties (their being a statue/a piece of clay) determined irrespective of what other properties they have at other times in the future. In fact, it is clear that this also means that sortal properties of objects are already determined at the very moment they start to exist. How this is accomplished depends on the kind of object we are considering. For example, for statues, it seems plausible to think that their sortal property is determined when they are created in a certain appropriate way that, I would say (but this is to be determined by the experts) includes the existence of a certain kind of person with the right kind of intentions in the right kind of environment and who makes the right kind of actions in relation to the right kind of amount of material. In the case of pieces of clay, their sortal property is determined when they are created in a certain appropriate way that, I would say, includes an amount of clay having its parts interconnected in the right kind of way. Our case is one in which these two sets of conditions obtain and therefore, from the very beginning it is determined that there is a statue, Goliath, and a piece of clay, Piece. Now, this idea can be combined with the idea belonging to the notion of object that, I would say, endurantism is trying to characterize that the sortal component of objects is of an excluding nature, that is to say, that objects that have sortal properties with existence and persistence conditions that may not be compatible are not the same object. In terms of our example, that objects that are statues cannot be pieces of matter (in the same sense as they cannot be tigers, coconuts or trees) and the other way around (of course) since, as we have seen, their persistence conditions may not be compatible. Remember that here I am not justifying the endurantist notion of object, but rather developing some of its features. Now, with the two ideas mentioned so far, we could explain how it is that in our situation we have a statue and a different object that is a piece of clay.

But then, how do we explain that being two different objects, they share so many properties? The feature of the endurantist notion of object that seems to explain this is the idea that objects have a second kind of component, a material component, which can be shared by different objects. This shared component would be the source of the shared properties. Moreover, these two components, the material component and the sortal component, are thought to be, on an

endurantist framework, as it happens on a perdurantist framework, of a mutually irreducible nature[3].

In short, a notion of object with the following features could explain the situation that results from the denial of the intuition behind the temporal grounding problem (denial that endurantists share with perdurantists): 1) objects have components of a dual nature: a *sortal*-component and a material component, 2) the sortal component is already determined at the very beginning of the object's life irrespective of the object's own futural properties, and 3) it is of an excluding nature. Furthermore, 4) the two components are mutually irreducible, and 5) the material component can be shared by different objects.

Two important remarks are in order here. First, the ideas above are quite bare bones ideas and so can be developed in different more specific ways. See [Paul (2006)] or [Koslicki (2008)].

Second, that these ideas are quite bare bones ideas does not mean that they do not constitute by themselves an endurantist answer to the temporal grounding problem. In any case, they are not less precise than the ideas that constitute the perdurantist solution to the temporal grounding problem.

Conclusion

I began the paper by presenting the temporal grounding problem for (a version of) perdurantism and (a version of) endurantism. As we saw, these two frameworks reject the intuition behind the temporal grounding problem and hold that in cases like the one of Goliath and Piece, there are two different objects which share some of their properties but differ in other properties that are irreducible to the first ones. In this paper, I have tried to emphasize that these two frameworks have in their notions of object enough resources to give an equally adequate and coherent explanation of the resultant situation. However, I have also emphasized that their resources are very different in nature.

This being so, and against what has usually been thought, the temporal grounding problem does not allow us to choose perdurantism over endurantism (nor the other way around, I would add).

References

Hawley, K. (2008), 'Persistence and determination', Philosophy 83 supplement 62, pp. 197-212.

Koslicki, K. (2008), The Structure of Objects, Oxford, Oxford University Press.

Paul, L. (2006), 'Coincidence as overlap', Noûs 40, pp. 623-59.

[3] In the perdurantist case the sortal property of an object is not determined just by the material properties of its temporal parts but by them along with the counterpart relations that unite its temporal parts.

Demonstrating fictional names*

Gemma Celestino Fernández
University of British Columbia & LOGOS
gceles@interchange.ubc.ca

Introduction

In this paper I want to propose a theory about the semantics of the fictional discourse that follows a particular thesis about the meaning of proper names, and singular terms in general. This proposal partly consists in an account of descriptive phrases in general such as 'the Prime Minister Gordon Brown' or 'our son John' as well. In particular, on this proposal fictional names do not refer to anything at all, but we manage to refer to fictional characters or to make true claims allegedly about the fictional persona they represent by using them as demonstrations within descriptive phrases such as the ones above.

The thesis about the semantics of singular terms is that in addition to referring, as most singular terms do, they also express a token-reflexive rule[1]. For our purposes, it is sufficient to say that for any token M of a proper name N, M semantically expresses the meaning of a rigid definite description such as 'the individual called M' –where M is the token actually used of the name[2]. Whereas the view I will propose is a descriptivist one, it is, as we will see, relevantly different from the descriptivist views on fiction of David Lewis (1983) and Gregory Currie (1988, 2003).

Fictional Discourse

The proposal is as follows.

Starting Point: Fictional Names and Sentences

Fictional names are empty rigid proper names: do not refer, are not abbreviations of any kind of definite description, and are not used as if they were abbreviated definite descriptions either. In telling fictions such as Tolstoy's *Anna Karenina*,

* I would like to specially thank Manuel Garcia-Carpintero, Dominic Lopes, Josep Macià, Genoveva Martí, Jeff Pelletier, Stefano Predelli, Pablo Rychter, Ori Simchen, and Steve Yablo for their helpful discussions and comments.

[1] Different versions of the token-reflexive view have been developed and defended by philosophers such as Hans Reichenbach, John Perry (1993) and Manuel García-Carpintero (1998, 2000).

[2] I take the rigidity comes from the fact that here it is the uttered token of the name that is mentioned, as I assume the referential properties of proper names are essential to them. And, hence, that it is not possible for a token of a name to exist at other possible worlds with different referential properties. However, if one does not agree with that assumption, a description such as 'the individual actually named M' will do.

fictional names are used as if, in pretence, they were non-empty rigid proper names that actually referred. As explained before, fictional names have meaning due to the meaning of the rigid definite description they, like any other proper name, express.

Further, not only simple linguistic expressions such as proper names perform more than one semantic function. In general, utterances of sentences express more than one proposition: a general descriptive proposition in addition to the primarily expressed one -the latter being the one that determines their truth-value. The meaning of those rigid definite descriptions expressed by fictional names constitute part of the general descriptive proposition secondarily expressed by utterances of sentences containing them.

Positive Statements in Fiction

Among the sentences with empty names, there are sentences with fictional names such as

(1) Anna Karenina suffered

which, like the other sentences with empty names, are meaningful but truth-valueless. For there is no one referred to by 'Anna Karenina', and indeed it could not be, due to the manifested intentions involved in using that fictional name as fictional and so, with no intention to refer. The sincere, literal and naïve use of these sentences would be made by confused children or other existence-of-fictions believers.

Apart from these, however, there are things that we say about, or within, the fiction that are true or false about, or in, the fiction. As it is widely assumed, to say things within the fiction that are true or false in a particular fiction, we use an operator which is specific to the case of fiction. For example, we express the truth that Anna Karenina suffered according to the fictional story that Tolstoy created, *Anna Karenina*, by uttering something like

(2) According to Tolstoy's *Anna Karenina*, Anna Karenina suffered

or some variant of it constructed with other similar phrases such as 'in the fiction'. (2) is true whereas (1) is truth-valueless. Sometimes, though, we express what utterances of (2) express without making the fiction operator explicit, that is, uttering the same sound pattern corresponding to (1). But it is clear that in these cases, the relevant expressions 'in the fiction' or 'according to the fiction' are implicitly used.

Part of the proposal is that what the fiction operator does is to change the interpretation of the embedded sentences by making the proposition these sentences normally only secondarily express be their primary content. Thus, an utterance of (2) primarily expresses

(3) According to Tolstoy's *Anna Karenina*, the individual called 'Anna Karenina' suffered

which is quite the same as saying

(4) that according to Tolstoy's *Anna Karenina*, there is an individual and only one who is called 'Anna Karenina' and suffered

(3) and (4) are true and explain the content and truth of (2). Again, it is important to bear in mind that it is the token used in an utterance of (2) that it is mentioned in (3) and (4). Hence, (3) and (4) involve rigid definite descriptions rather than non-rigid ones.

It is precisely this rigidity that accounts for the specificity of fictional stories explained by the use of singular terms such as Tolstoy's *Anna Karenina* and many others. Unlike Lewis and Currie's view, the current proposal is that even if fictional stories cannot be singular due to their being fictional, they can yet be specific rather than general this way: by means of using *rigid* singular terms. Thus, on the general view I am arguing for, there are three grades ranging between generality and particularity that may be expressed through language: the generality expressed by the use of non-rigid definite descriptions, the singularity expressed by the use of non-empty singular terms and the specificity that lies in between and that can be expressed by the use of rigid definite descriptions as the ones that are conventionally associated with proper names, among others. Further, the rigidity involved in the fictional discourse entails that most fictional stories and the fictional persona they represent are not even possible in an important sense: there is no possible thing satisfying these descriptions.

Now. In addition to these fictional sentences, there are also statements we use to talk about fiction such as

(5) Sherlock Holmes is a fictional character

(6) Conan Doyle created Sherlock Holmes

These statements do not involve a prefixing of the fiction operator like the statements made to talk about the things that fictionally happen in a given fiction we considered above.

(7) In the fiction, Sherlock Holmes is a fictional character

(8) In the fiction, Conan Doyle created Sherlock Holmes

(7) and (8) are clearly false and not what utterances of (5) and (6) express, which seem to be true, or have one true reading.

My proposal partly consists in extending the hypothesis about the existence and prefixing of the fiction operator 'in the fiction' or 'according to the fiction' to other fictional expressions such as 'the fictional character' and 'the fictional persona', so that the contents of utterances of (5) and (6) are, respectively, something like the following

(9) The fictional character, Sherlock Holmes, is a fictional character

(10) Conan Doyle created the fictional character Sherlock Holmes

These other fictional expressions will make it salient that the fictional name in these statements is not used to pick out a referent, but is merely demonstrated to identify and thereby help to finally make reference with the whole descriptive complex expression to the fictional character in question, or no reference at all in

the cases of 'fictional persona' and the like[3]. This prefixed fictional expression does this by making it clear that the statement is about fiction. The idea is to interpret (9) and (10) somehow analogously as we interpreted (2) and (3). In (5), (6), (9) and (10), the expression 'the fictional character' is prefixed and the name 'Sherlock Holmes' is in a way "quoted" and, hence, demonstrated, rather than used, with the sole purpose of identifying the particular fictional character the statement is about. 'Sherlock Holmes' does not refer to that fictional character either; the way we get to talk about, and hence refer to, the fictional character is by using the whole complex expression 'the fictional character, Sherlock Holmes'.

My proposal is that the primary content expressed by (5) and, in turn by (9), is something like the content that would be primarily expressed by an utterance of

> (11) There is one and only one fictional character, this, and is a fictional character - where 'this' refers to the existing abstract fictional character Sherlock Holmes and contributes it as a component to the proposition.

This proposition is expressed because of a relation to the fictional name that is made contextually salient by means of a demonstration of the name and expression of the meaning of the description in question. (5) and (9) may be characterized as obtaining their primary proposition by saying something such as the following

> (12) The x such that x is a fictional character and R (x, 'Sherlock Holmes'), x=M, is a fictional character - where R stands for a contextually salient relation between the fictional character and the fictional name and M stands for a name of the fictional character.

These uses of complex expressions which combine proper names and descriptions not only exist in the fictional discourse. These are the cases of ordinary descriptive phrases such as 'the Prime Minister Gordon Brown' or 'our son John', for instance. So, in general, the proposal is that for any descriptive phrase containing a proper name like these, DN, the name N is used as a demonstration to pick out the individual satisfying the descriptive phrase DN in question: that is, satisfying the descriptive part D and bearing the contextually relevant relation to the name N. This contextually salient relation is in many cases just the one of being referred to by the name.

Finally, the involvement of fictional expressions like 'the fictional character' or 'the fictional persona' in the fictional discourse is quite abundant and flexible. Other sentences such as, for instance,

> (13) Sherlock Holmes is smarter than Poirot

might also involve the implicit use of expressions similar to them, but different in certain respects. (13) perhaps expresses something like

---

[3] I am making a substantial metaphysical assumption which I think is true: that there is a distinction between fictional characters and the fictional persona whose existence they merely represent. In short, fictional characters are abstract objects that therefore exist. But fictional persona allegedly represented by fictional characters, would in many cases be non-existent concrete individuals such as you and me, but do not exist.

(14) The fictional detective, Sherlock Holmes, is fictionally smarter than the fictional detective Poirot

All of these fictional phrases have in common that they use the expression 'fiction' plus something else to make the point or advertisement that one is talking about fiction.

Negative Singular Existentials in Fiction

There are yet other uses of sentences about fiction as such or from the outside in need of explanation. These are the true fictional negative singular existentials. Sentences such as

(15) Santa Claus does not exist

These seem to be prefixed by an indefinite description rather than a definite one:

(16) A fictional persona, Santa Claus, does not exist

with the following true reading

(17) It is not the case that there is a fictional persona Santa Claus that exists

which would directly get the content intended, i.e.

(18) No fictional persona Santa Claus exists

In cases where there is nothing that satisfies the description and contextual relation to the name nothing more than what the description expresses is contributed to the primary proposition expressed by the utterance of the sentence. True negative singular existentials such as (15) are cases like these.

In cases where there is nothing that satisfies the description and bears the contextual relation to the name, the contextual relation is contributed to the proposition instead of an individual. Thus, 'A fictional persona Santa Claus', for instance, contributes something like

(19) An x such that x is a fictional persona and R (x, 'Santa Claus') -where R stands for the relation of being represented by or being associated to.

The proposal would apply analogously to other empty names that are not fictional such as 'Vulcan'. For instance, sentences in which the name 'Vulcan' occurs are sometimes prefixed (be it explicitly or implicitly) by a descriptive phrase such as 'the hypothesized planet'.

References

Currie, G. (2003), 'Characters and Contingency', *Dialectica* 57, pp. 137-48.

García-Carpintero, M. (2000), 'Indexicals as Token-Reflexives', *Mind* 107 (1998), pp. 529-63.

Lewis, D. (1983), 'Truth in Fiction', in Lewis, D., *Philosophical Papers* 1, Oxford, Oxford University Press.

Perry, J. (1993), *The Problem of the Essential Indexical and Other Essays*, Oxford, Oxford University Press.

Causalidad mental y autoconocimiento

Flor Emilce Cely Ávila
Universidad Nacional de Colombia
ecelyf@unal.edu.co

El propósito de este artículo es doble: por un lado, analizar el problema de la ineficacia causal de los estados mentales en la teoría causal de la acción y, por otro, proponer un modelo de explicación de la misma que tenga en cuenta como aspecto esencial los rasgos especiales del autoconocimiento. El punto de partida será la explicación causal de la acción propuesta por Davidson; como es sabido, este autor hizo una crítica a ciertas concepciones de la explicación por razones de la acción en las que no se esclarecía la naturaleza de la relación entre las razones y la acción; pues, señalaba, uno puede tener razones para hacer algo y hacerlo y, sin embargo, no realizar la acción por esas razones. Al no contar con una explicación satisfactoria del tipo de conexión que hay entre una acción y las razones que la explicarían, un tipo de explicación anti-causalista hace que tal conexión resulte "misteriosa". Es por esta razón que Davidson plantea que las racionalizaciones o explicaciones por razones deben ser consideradas como explicaciones causales y a las razones como causas de la acción [Davidson (1963)]. Esto en el marco de su tesis más amplia del monismo anómalo, que defiende estas tres premisas: (1) Los estados mentales se relacionan *causalmente* con estados físicos; (2) Las relaciones causales singulares caen bajo leyes deterministas estrictas; y (3) No hay leyes psicológicas ni psicofísicas estrictas [Davidson (1970)].

Ahora bien, uno de los mayores problemas que enfrenta esta concepción causal tiene que ver con el hecho de que en ésta se dejaría a los estados mentales como causalmente ineficaces. Varios autores han señalado que esta posición monista no reduccionista se enfrenta a la acusación o bien de inconsistencia, o bien de epifenomenismo [*cfr*., por ej. Crane y Brewer (1995); y McDonald y McDonald (1995)]. Por un lado, sería *inconsistente*, dado que si los estados mentales producen efectos en virtud de sus propiedades mentales, entonces debería haber leyes psicofísicas, y el carácter anómalo de lo mental excluye esta posibilidad. Y, por otro, si los estados mentales tienen tales efectos en virtud de sus propiedades físicas, entonces debe enfrentar el cargo de *epifenomenismo*; esto es, defender una posición en la que la causalidad de los procesos mentales opera sólo gracias a que son idénticos a procesos físicos y en la que se opta por una concepción nomológica de la causalidad, desemboca en la idea de que no hay realmente eficacia causal de lo mental. Pues de esta manera las únicas propiedades causalmente relevantes serían las propiedades físicas y lo que interesa entender es cómo un evento mental produjo, llevó a cabo, o causó una acción en virtud de sus propiedades mentales (esto es, en virtud de su contenido) y entender esto es lo que vale la pena cuando se habla de "causación mental". Parece ser entonces que, a pesar de los esfuerzos de Davidson por defender una idea de causalidad mental

[*cfr*. Davidson 1995], al final no lo logra, pues está comprometido con una concepción de la causalidad que no da luces sobre la manera cómo los estados mentales en *cuanto tales* tendrían realmente eficacia causal en la acción [*cfr*. Antony (1989) y Kim (1995)].

Ante las dificultades que implica para la teoría causal de la acción la acusación de epifenomenismo propongo considerar el papel fundamental que juega el autoconocimiento en la determinación de las razones para actuar y, por tanto, en el entendimiento y explicación de la acción por parte del agente mismo y de los otros. Recordemos que en la propuesta de Davidson las razones son consideradas como causas de la acción y las racionalizaciones como explicaciones causales. Pero lo importante aquí es que Davidson está interesado en defender una concepción causal de la acción que parte del reconocimiento del papel de justificación racional de la explicación por razones y de la conexión lógica que hay entre las razones y la acción (algo que, por otra parte no excluye que tal relación sea causal). Ahora bien, considero que una manera apropiada de darle fuerza a este papel de justificación consiste en otorgarle a la perspectiva y autoridad de primera persona (que tiene el agente sobre sus propios estados mentales) un papel explicativo fundamental. En este sentido interesa hacer énfasis no sólo en que el papel causal del agente no es incompatible con la perspectiva de primera persona, sino en que en una teoría causal como la que Davidson pretende defender es necesario concederle un rol explicativo a esta perspectiva personal. La idea aquí entonces es que la fuerza explicativa de las razones viene dada por el carácter especial del autoconocimiento de nuestras razones para actuar.

Se ha afirmado que el conocimiento que tenemos de nuestros propios estados mentales es especial en tanto (i) tiene un carácter inmediato o no inferencial; (ii) es independiente de la observación empírica; y (iii) goza de una especial presunción de verdad y resistencia al error. En general, estos rasgos caracterizan lo que se conoce como autoridad de primera persona. A pesar de que algunos afirman que las razones para actuar no son transparentes para nosotros en la medida en que lo son nuestras sensaciones y sentimientos, interesa aquí defender la idea de que cada uno de nosotros generalmente es autoridad respecto a las razones para actuar. Así, necesariamente se debe contar con un agente racional que *esté en capacidad*, o que *le sea posible*, dar cuenta de sus acciones, articulando sus razones desde una perspectiva de primera persona, pues será esto lo que nos permitirá llegar a la explicación correcta de la acción (identificando la razón que fue su causa), sin dejar por fuera la justificación de la misma.

Ahora bien, esta afirmación no implica que *todo el tiempo* tengamos un conocimiento transparente de nuestras propias razones o que seamos una autoridad *infalible* respecto a las mismas (pues ciertamente hay excepciones, como cuando nos declaramos confundidos respecto a las razones que nos llevaron a actuar, o cuando nos vemos en la posición de hacer inferencias o sacar conclusiones de nuestras razones para actuar basados en evidencias). Pero, aunque la posibilidad de equivocarse está siempre presente, no se debe negar que la mayoría del tiempo tenemos una autoridad especial sobre las creencias y deseos que constituyen las razones que nos llevan a actuar y que no necesitamos de la observación, evidencia

o de una perspectiva objetiva para dar cuenta de las mismas. Lo que interesa resaltar aquí es que este rasgo especial del autoconocimiento de las razones para actuar es un dato esencial en la explicación de la acción, si es que quiere darse una explicación que nos muestre las razones por las cuales el agente efectivamente llevó a cabo la acción. De hecho, podríamos afirmar que si los casos que hemos considerado como excepciones (casos en los que una persona no tienen ni idea de lo que está haciendo o por qué) se convierten para alguien en algo común, difícilmente podríamos considerarlo un agente.

Según Kim (1998), un agente entiende su acción en la medida en que conozca la razón primaria sobre cuya base escoge, o podría escoger si hubiera deliberado, hacer lo que hizo. Esto no implica que todo el tiempo escojamos o deliberemos conscientemente sobre lo que hacemos, pues muchas de nuestras acciones son llevadas a cabo más o menos automáticamente y sólo *ex post facto* reconstruimos la razón primaria. Sin embargo, nuestra habilidad de hacerlo es esencial para el auto-entendimiento como agentes reflexivos:

> El auto-entendimiento surge del contexto de la deliberación, elección y decisión. El contexto de la deliberación es necesariamente un contexto de primera persona. Pues cuando deliberas debes nombrar lo que quieres y deseas sobre el mundo desde tu perspectiva interna y *esa es la única cosa que puedes nombrar*. Las bases de tu deliberación deben ser internamente accesibles […] Las razones para actuar, entonces, son necesariamente *razones internas*, razones que son cognitivamente accesibles para el agente. Este es un tema crucial en el cual las razones para actuar difieren de las causas de las acciones: las razones deben, aunque las causas no necesitan ser, accesibles para el agente". (Kim *ibid*. pp. 78).

En general, estoy de acuerdo con el énfasis puesto por este autor en que la posibilidad de dar las razones de la acción tiene que ver con la perspectiva de primera persona del agente; sin embargo, considero que hay un problema con la diferencia que plantea al final del párrafo entre razones y causas, es decir, con la idea de que, a diferencia de las razones, las causas que utilizamos para explicar la acción no necesitan ser accesibles para el agente. Esto tiene que ver con una dicotomía –aceptada por muchos autores– que asocia, de un lado, una perspectiva impersonal con la explicación causal y, de otro, la perspectiva de primera persona, con la comprensión, la interpretación o la explicación no causal de la acción.

La propuesta es superar estas dicotomías y entender que la explicación causal de la acción, no sólo no es incompatible con una perspectiva personal, sino que dicha perspectiva es fundamental para llegar a una explicación correcta de la acción. Pues, mientras que las razones para actuar son designadas teniendo en consideración al agente, su perspectiva personal, en el caso de los demás eventos físicos podemos designar las causas desde una perspectiva completamente impersonal (como en el caso de indagar por las causas del colapso de un puente en el que, por supuesto, no se necesita contar con la 'perspectiva o punto de vista del puente').

En síntesis, a diferencia de las corrientes anti-causalistas [*cfr*. por ejemplo, Ginet (1990)], no considero necesario renunciar a la idea de una explicación

causal de la acción. Pero haciendo énfasis, asimismo, en que es imprescindible en este tipo de explicación tener en cuenta como fundamental el papel explicativo que juega la perspectiva personal. Así, el elemento que haría falta a la explicación causal de la acción, más allá del complejo o la fuerza motivacional que proponen sus críticos [*cfr*. Tanney (1995) y Dickenson (2007)], tiene que ver con la perspectiva de primera persona a partir de la cual el agente, en la mayoría de casos, puede articular las razones que lo llevaron a actuar con una autoridad epistémica de la que carece una perspectiva impersonal y puede llegar, de esta manera, a seleccionar cuál fue *la* razón que lo llevó a actuar y, al mismo tiempo, a descartar las racionalizaciones alternativas.

La perspectiva personal se considera entonces de una importancia fundamental en la explicación causal de la acción, y no sólo en el caso de las acciones propias, sino en las posibilidades de entendimiento de las acciones de los otros cuando nos situamos en una perspectiva, bien sea de segunda o de tercera persona. Consideremos, por ejemplo, cómo designamos *la* razón de alguien para actuar. Generalmente lo hacemos preguntando a la persona y aceptando su respuesta, pues tendemos a tratar a otros como *autoridad* respecto a sus acciones. Esto es, la razón para actuar debe ser designada desde el 'punto de vista personal': de alguna manera cada uno en su propio caso decide que un par deseo/creencia tiene prioridad explicativa sobre otros y entonces lo designa como su causa. En este sentido, podemos entender así la idea de racionalización:

> … una razón que racionaliza una acción tiene que ver con ser la causa de esa acción, y explica la acción (en parte) revelando algo que el agente tenía el propósito de llevar a cabo y, de esta manera, *algo que hace la acción "razonable" o "aceptable" en algún grado, desde el punto de vista del agente*. Obviamente, la racionalidad asociada con la racionalización es entendida de una manera fina y subjetiva… (Mele 2003, pp. 71, las cursivas son mías).

Referencias bibliográficas

Antony, L. (1989). 'Anomalous Monism and the Problem of Explanatory Force', *The Philosophical Review* 98, pp. 153-87.

Crane, T. y Brewer, B. (1995), 'Mental Causation', *Proceedings of the Aristotelian Society*, Supplementary Volumes 69, pp. 211-53.

Davidson, D. (1963) [1995], 'Acciones, razones y causas', en Davidson, D. *Ensayos sobre acciones y sucesos*, O. Hansberg, J. Robles y M.Valdés (trad.), UNAM, Barcelona, Crítica, pp. 17-36.

Davidson, D. (1970) [1995], 'Sucesos mentales', en Davidson, D. *Ensayos sobre acciones y sucesos*, O. Hansberg, J. Robles y M.Valdés (trad.), UNAM, Barcelona, Crítica, pp. 263-87.

Davidson, D. (1984) [2003], 'La autoridad de primera persona', en Davidson, D., *Subjetivo, intersubjetivo, objetivo*, O. Fernández (trad.), Madrid, Cátedra, pp. 25-40.

Davidson, D. (1995), 'Thinking Causes', en Heil, J. y Mele, A. (eds.), *Mental Causation*, Oxford, Clarendon Press, pp. 3-17.

Dickenson, J. (2007), 'Reasons, Causes, And Contrasts', *Pacific Philosophical Quarterly* 88, pp. 1–23.
Ginet, C. (1990), *On Action*, New York, Cambridge University Press.
Heil, J. y Mele, A. (eds.) (1995), *Mental Causation*, Oxford, Clarendon Press.
Kim, J. (1995), 'Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?', en Heil, J. y Mele, A. (eds.), *Mental Causation*, Oxford, Clarendon Press, pp. 19-26.
Kim, J. (1998), 'Reasons and the First Person', en Brensen, J. y Cuypers, S. (eds.), *Human Action, Deliberation and Causation*, Philosophical Studies Series 77, Dordrecht, Kluwer, pp. 67-87.
Macdonald, C. y Macdonald, G. (1995), 'How to Be Psychologically Relevant?', en McDonald, C. y McDonald, G. (eds.), *Philosophy of Psychology*. *Debates on Psychological Explanation*, Londres, Backwell, pp. 60-77.
Mele, A. (2003), *Motivation and Agency*, New York, Oxford University Press.
Tanney, J. (1995), 'Why Reason May Not Be Causes', *Mind and Language* 10, pp.105-28.

Why a psychologist doesn't need to be a constructivist

Antonella Corradini
Catholic University of Milan
antonella.corradini@unicatt.it

Introduction

Social constructionism presents a radical challenge to cognitivistic psychology. It rejects the methodological requirements which characterize the latter, such as experimental method and laboratory research, and denies that psychology should be considered as one of the natural sciences. The social constructionistic movement's aim is to renew categories and concepts, such as intentionality, meaning, agency, relationality, which have been neglected by the cognitivistic psychology, and to claim the main role of psychology as representative of the *Geisteswissenschaften*.

However, to claim this role – according to constructionism – is tantamount to take a turn from a realistic to a constructivistic view of mind and reality. (The term "constructionism" is preferred to "constructivism" to stress that mind and reality are the products of social interaction.) What reasons can be put forward in favour of this latter thesis? The answer given by the present paper is: none! I'll argue in favour of my view in three steps.

1. I'll first analyse how the turn from realism to constructivism occurred in early modern philosophy, taking as examples John Locke's realism and George Berkeley's immaterialism. Although Locke assumes that the ideas of the primary qualities – in contrast with the ideas of the secondary qualities – do correspond to the reality of things, he gives a general definition of knowledge which seems incompatible with the previous view: knowledge consists in the perception of the concordance or non-concordance with the world of the ideas we possess. According to this definition, our knowledge consists in nothing more than the possession of ideas. " … it seems probable to me, that the simple ideas we receive from sensation and reflection, are the boundaries of our thoughts; beyond which the mind, whatever efforts it would make, is not able to advance one jot; nor can it make any discoveries, when it would pry into the nature and hidden causes of those ideas." [1690, II, XXIII, §29] Locke's realism is a form of mediate realism, according to which we do not know things themselves, but only our representations of things, that is to say the ideas.

George Berkeley was aware that mediate realism can lead to scepticism: if knowledge is confined only to ideas, there cannot be a knowledge of material objects, because we cannot compare the ideas with something that is not an idea, that is to say the external world that produces these ideas. Berkeley's move for

defeating scepticism consists in denying the existence of a material substance: if a material substance does not exist, the nature of things consists in their being perceived. "For as to what is said of the absolute existence of unthinking things without any relation to their being perceived, that seems perfectly unintelligible. Their esse is percipi, nor is it possible they should have any existence out of the mind or thinking things which perceive them." [1710, § 3] Berkeley's philosophy, immaterialism, is a kind of constructivism, made necessary by the urge to escape from scepticism. However, it is worth noting that the passage from realism to constructivism is not due to realism itself but to a specific kind of realism, i.e *mediate* realism.

2. In the second part of this paper I'll examine how realism is understood by two contemporary psychologists, George Kelly and Jerome Bruner, who both give their allegiance to constructivism (as well as to constructionism, in Bruner's case).

In present-day psychology, constructivism mainly owes its popularity to the (often unquestioned) acceptance of pragmatism and of this latter´s tendency to substitute the realistic concept of truth with the instrumentalistic notion of utility [Kelly (1955), 1, p. 17; Bruner (1990), p. 24 ff]. Kelly, for example, tells us that "… there are various ways in which the world is construed. Some of them are undoubtedly better than others. They are better from our human point of view because they support more precise and more accurate predictions about more events" [Kelly (1955), 1, pp. 14-15]. However, we could ask ourselves why, by telling this, Kelly deems it necessary to embrace a "constructive alternativism" and to abandon realism. As a matter of fact, the validity of predictions could be interpreted as a hint of the truth of the theory from which they derive. Kelly answers the previous question in the following, somehow surprising way: "…since we insist that man can erect his own alternative approaches to reality, we are out of line with traditional *realism*, which insists that he is always the victim of his circumstances [Kelly (1955), 1, p. 17]. The reason why an alternativistic view must be constructive is, then, that realism allows for no plurality, since it only admits of a unique, true way of describing the world.

Jerome Bruner also champions a position similar to Kelly's as regards the pluralism of viewpoints about reality. A sympathiser with Nelson Goodman's philosophy of science, Bruner maintains in fact that the manifold possible worlds produced by human beings do not need to correspond to a real and objective world in order to be valid. "…it is far more important, for appreciating the human condition, to understand the ways human beings construct their worlds … than it is to establish the ontological status of the products of these processes. [Bruner (1985), p. 46]: "…no one "world" is more "real" than all others, none is ontologically privileged as the unique real world." [Ibidem, p. 96]

Once again, however, we could plausibly ask ourselves why Bruner's pluralism of possible worlds should be incompatible with realism. Bruner's antipathies for realism are likely to arise from the neo-positivistic identification of reality with physical reality, with the consequent exclusion of the subjective dimension of the mental from the scientific domain. In Bruner's view, the

rejection of physicalism would exempt psychology from meeting the requirements of the natural sciences, considered by the neo-positivists as the only objective ones. From the constructive point of view, instead, both the physical and the mental domains are the results of different constructive activities, none of them is allowed to have an ontological preminence over the other.

The frequent mention of the Vienna Circle's physicalism, however, should lead us to surmise that Bruner tends to identify realism with a specific kind of realism, causal realism. According to this view, a causal relation holds between mind and reality: stimuli are transmitted from the external objects to the subject's brain which successively elaborate them. If this is the kind of realism Bruner and Kelly refer to, then many of their claims can be seen in a different light. As an example, Kelly's apparently bizzarre thesis that realism compels human beings to passivity is a result both of causal realism itself and of the – more or less proper – way behaviourism makes use of it. A clear example of the passive role to which the agent would be forced in the frame of a behaviouristic perspective is illustrated by Skinner's "Selection by Consequences": "The role of selection by consequences has been particularly resisted because there is no place for the initiating agent … A proper recognition of the selective action of the environment means a change in our conception of the origin of behavior … so long as we cling to the view that a person is an initiating doer, actor, or causer of behavior, we shall probably continue to neglect the conditions which must be changed if we are to solve our problems" [Skinner (1981), p. 504].

As regards Bruner, the point of his severe criticism of the cognitive revolution becomes clearer if we reflect upon the epistemological affinity of behaviourism with the cognitive sciences. It is undeniable that these latter give a more active and creative image of the human being as elaborator, through mental processes, of the physical input into the behavioural output. Nevertheless, the difference between behaviourism and cognitive sciences is only one of degree, and not a qualitative one, as it is shown by their common sharing of both physicalism and causal realism. Not by chance is it that an authoritative cognitivistic philosopher such as Jerry Fodor does not eschew from declaring – speaking about intentional content – "that *dog* thoughts are about dogs because they are the kinds of thoughts that dogs can be relied upon to cause. Similarly, mutatis mutandis, for thoughts with other than canine contents" [Fodor (1994), pp. 4-5].

We can therefore concede that Kelly and Bruner have good reasons for preferring constructivism to realism. However, it must be stressed that also in this case the turn to constructivism is not grounded in realism itself but arises from a reaction to a special kind of realism, *causal* realism, which is bound to radical empiricism.

3. As a consequence of my reflections in sections 1 and 2, I surmise that the necessity to move from realism to constructivism in order to safeguard the autonomy of psychology as a social science is only an apparent one. Constructivism, actually, is needed only under the condition of ignoring *intentional* realism. This latter conceives knowledge as the presence of the object to the subject as something other than the subject herself. This form of realism is

different from mediate realism since, according to it, the mind has got a direct relation to reality as intentional object of the knowing act. On the other hand, it is different from causal realism as well, because in the context of intentional realism the relation with reality is viewed not as external, natural and concrete but as internal, ideal and abstract.

It can be shown that adoption of intentional realism allows us to block the move from realism to costructivism both in the first and in the second case under scrutiny. As far as the first case is concerned, intentional realism does not even let the worry arise that induced Berkeley to choose immaterialism to escape scepticism. There is no question of comparing our representations with reality, because, according to intentional realism, our knowing acts are directed from the very beginning towards reality. To be sure, access to reality is impossibile without any representations of it, but knowledge must not be restricted within the limits of our representations. To put it in a nutshell, while according to mediate realism we only know our representations of reality, according to intentional realism we know reality through our representations of it. This latter point explains well why, in the second case illustrated above, the adoption of intentional realism makes costructivism dispensable. The directedness of the mind towards reality, in fact, goes along with its perspectivity. Every knower approaches reality from a specific point of view, which is peculiar to her and/or the species she belongs. Pace Kelly, to be a realist does not amount to being a victim of the circumstances in which we live. Intentional realism, in fact, is able to account for the many and diverse ways in which human beings can attain knowledge of reality [on intentionality cf. Crane, (2001), ch. 1; Oderberg, (2008)].

My conclusion is that intentional realism is perfectly compatible with the status of psychology as a social science. I would also like to argue that it accounts for such a status better than constructivism/constructionism itself, but lack of space obliges me to do this in another essay.

References

Berkeley, G. (1710), *A Treatise Concerning the Principles of Human Understanding.*

Bruner, J. (1985), *Actual Minds, Possible Worlds*, Harvard University Press.

— (1990), *Acts of Meaning*, Harvard University Press.

Crane, T. (2001), *Elements of Mind. An Introduction to the Philosophy of Mind*, Oxford, Oxford University Press.

Fodor, J. A. (1994), *The Elm and the Expert. Mentalese and its Semantic*, Cambridge (MA), A Bradford Book, the MIT Press.

Kelly, G. A. (1955), *The Psychology of Personal Constructs*, 2 vol., New York, W. W. Norton & Company Inc.

Locke, J. (1690), *An Essay Concerning Human Understanding.*

Oderberg, D. S. (2008), 'Concepts, Dualism, and the Human Intellect', in A. Antonietti, A. Corradini, and E. J. Lowe (eds.), *Psycho-physical Dualism Today. An Interdisciplinary Approach*, London-New York, Lexington Books, a division of Rowman and Littlefield Publishers, pp. 211-233.

Skinner, B. (1991), 'Selection by Consequences', *Science* 213, pp. 501-504.

Compromisos individuales y compromisos sociales: el problema del reduccionismo

Miranda del Corral de Felipe
Universidad Nacional de Educación a Distancia
mdelcorral@bec.uned.es

El concepto de compromiso ha experimentado un uso creciente tanto en ciencias sociales, especialmente en economía y en psicología social, como en filosofía de la mente y de la acción. Sin embargo, no es frecuente la profundización o el debate acerca de la naturaleza de este concepto, es decir, a qué acciones o estados mentales representa o describe. A nivel metodológico, consideramos adecuado plantear el problema del reduccionismo entre el compromiso social y el compromiso individual, ya que el mismo concepto es empleado tanto para referirse a estados mentales internos, como a acciones sociales que implican varios agentes.

El compromiso individual o interno

El compromiso puede ser entendido como un rasgo inherente a las intenciones: cuando un sujeto tiene la intención de realizar una acción, tiene una grado de compromiso mayor o menor con la meta que pretende alcanzar realizando esa acción (Bratman, 1997, 2004; Cohen y Levesque, 1990). Otros autores como Elster (1979, 2000) o Searle (2001) defienden que el compromiso sólo aparece cuando el sujeto prevé que la acción que pretende llevar a cabo está motivada por deseos variables en el tiempo y dependientes del contexto, y que en el momento de realizar la acción puede no tener el deseo de hacerla: el compromiso sería, por lo tanto, una razón para la acción independiente de los deseos. Sen (1977, 1985, 2005) defiende una concepción más restringida de compromiso, para referirse a acciones motivadas por normas y valores morales, y no sólo independiente de las preferencias del sujeto, sino contraria a ellas.

Por otra parte, Millar (2004) defiende que el compromiso interno no sólo afecta a intenciones (compromisos de acción), sino a creencias, prácticas y significados. En el caso de las intenciones, Millar afirma que adoptar una intención implica comprometerse a realizar aquello que sea necesario para alcanzar la meta. El carácter obligatorio de estas acciones conforman el aspecto normativo del compromiso.

Por lo tanto, el compromiso puede entenderse de dos maneras. La primera, como un evento mental que "causa" (motiva, precede) la acción, siendo un componente necesario de las intenciones, pudiendo ser independiente o contrario a los deseos. La segunda manera de entenderlo es como la consecuencia normativa de una intención (o de una creencia) que constriñe nuestro comportamiento futuro

para alcanzar una meta previa, generando un conjunto de acciones de carácter obligatorio.

El compromiso social

Otras teorías analizan la dimensión social del compromiso, es decir, entendiéndolo como fruto de la interacción entre al menos dos sujetos. Desde la teoría de los actos de habla (Austin 1962; Searle 1969), el compromiso es el resultado de un acto de habla en el que el sujeto adquiere la obligación de llevar a cabo una acción, como pueden ser las promesas y las amenazas. Dentro del ámbito de la teoría de juegos, Schelling (1960) propone el siguiente problema: si no confío en que los demás van a cooperar, ¿cómo puedo confiar en su compromiso de cooperar? Este autor propone otra definición de compromiso social más precisa (Schelling 1960, 2001): el compromiso es la asunción de la obligación de llevar a cabo un curso de acción, por medio de la restricción de las propias elecciones, con el fin de manipular el comportamiento de otra persona.

Castelfranchi (1996, 2003), al contrario que Schelling, defiende que la amenaza no es un tipo de compromiso social, sino una mera declaración de intenciones: si un sujeto no cumple su amenaza, no recibirá sanción por parte del sujeto amenazado. El compromiso, a diferencia de una declaración pública de intenciones, es generador de derechos y obligaciones.

De nuevo, podemos entender el compromiso social de dos formas. La primera sería la comunicación de intenciones a otro sujeto, o bien la asunción de una meta de otro sujeto como propia. En este caso, el compromiso (sea individual o social) sería la razón o el motivo para actuar por parte del sujeto que se compromete. La segunda forma de compromiso social es el acto social en el que un agente adquiere una obligación para con otro, y ambos son conscientes de ello. El compromiso sería la consecuencia normativa de una acción previa.

Compromisos individuales y sociales

Las cuestiones que nos planteamos aquí son las siguientes: ¿qué relación existe entre el compromiso individual y el compromiso social? ¿Es necesario (o suficiente) el primero para que se dé el segundo? De no ser así, ¿es adecuado referirnos con dos términos idéntico a fenómenos independientes?

De manera simplificada, existen cuatro posibles relaciones entre el compromiso individual y el compromiso social. Emplearemos el concepto de intención para referirnos a la precondición de existencia del compromiso individual, ya que todos los conceptos de compromiso individual que hemos analizado se refieren a éste para señalar un aspecto de la acción intencional (aunque los enfoques, como hemos visto, varían). Dependiendo de si el compromiso social es honesto o deshonesto, y de si la acción de realiza o no, obtenemos:

> a) X tiene la intención de hacer A; X se compromete con Y a hacer A; X hace A.

b) X tiene la intención de hacer A; X se compromete con Y a hacer A; X no hace A.

c) X no tiene la intención de hacer A; X se compromete con Y a hacer A; X no hace A.

d) X no tiene la intención de hacer A; X se compromete con Y a hacer A; X hace A.

Son problemáticos el caso c) y el caso d). Si X nunca ha tenido la intención de hacer A, ¿podemos afirmar que existe un compromiso por el hecho de que exista una comunicación verbal entre X e Y (una promesa, por ejemplo)? Autores como Castelfranchi o Millar afirman que el sí: la existencia de un compromiso social depende del acto de comunicación por el que se establece (sea honesto o deshonesto) y de la creencia de Y de que X está comprometido a hacer A. Para Schelling, la existencia de casos como b) y c) suponen, como hemos visto, el problema paradigmático de los compromisos: que no son una garantía de que X vaya a hacer a. Tampoco los análisis de la confianza en los compromisos sociales no suelen diferenciar entre los casos b) y c), puesto que en ambos casos el resultado es el mismo.

Creemos que la aparente contradicción entre que, por una parte, los compromisos sociales parezcan una continuación de los individuales, y que por otra parte, los compromisos sociales puedan existir sin los individuales, se basa en la confusión de dos aspectos del compromiso: el enfoque causal y las consecuencias normativas que se desprenden de la interacción humana. Dependiendo de dónde situemos el foco de análisis, podremos afirmar que el compromiso individual es suficiente, necesario o ni suficiente ni necesario para el compromiso social.

El compromiso como causa o motivación

Entendiendo el compromiso como el motivo de la acción, el compromiso individual es necesario, aunque no suficiente, para el compromiso social. En los casos clásicos a) y b), la motivación de X para intentar hacer A es el hecho de mantener un compromiso con Y de hacer A. Para que exista esa motivación, el compromiso ha de ser honesto, es decir, que cuando X promete a Y que hará A, ha de adoptar A como un compromiso interno. De no ser así, incurriríamos en situaciones donde X podría decir "te prometo que haré A, aunque no tengo ninguna intención de hacer A", lo cual resulta contradictorio. Ahora bien, ¿sigue existiendo un compromiso cuando X no tiene la intención de hacer aquello a lo que se compromete? Desde el punto de vista causal, no. Pero, ¿tiene derecho Y a penalizar a X por no haber sido sincero, o bien tiene derecho a penalizarlo por no haber cumplido el compromiso?

El compromiso como situación normativa

De no haber sido sincero, Y tiene derecho a castigar a X por no haber cumplido su compromiso, ya que el compromiso social es generador de derechos y deberes

desde el momento de su formulación. Pero, en este punto, es importante realizar una distinción: no son iguales las situaciones a) y d). De hecho, a) y d) tienen dos interpretaciones posibles:

> a, [d]1) X [no] tiene la intención de hacer A; X se compromete con Y a hacer A; X intencionalmente hace A.
>
> a, [d]2) X [no] tiene la intención de hacer A; X se compromete con Y a hacer A; X hace A accidentalmente, o con otro propósito ajeno al compromiso con Y.

Algunos autores, como Castelfranchi, sostienen que a2) y d2) no cumplen los requisitos para ser un compromiso satisfecho, ya que es condición necesaria que A se realice intencionalmente, al igual que ocurre en los compromisos individuales. Por ejemplo, si X promete a Y que apagará la televisión antes de acostarse, y se va la luz en ese momento, no puede decirse que X haya cumplido su promesa: decimos que las circunstancias han cambiado y el compromiso ya no es válido. Sin embargo, el caso de d1) es más controvertido. Hay una reconsideración por parte de X, y pasa de no tener una intención a tenerla. No trataremos este caso en este artículo, ya que no afecta al argumento principal.

La pregunta, entonces, sería la siguiente: ¿por qué son relevantes las intenciones de X cuando hace A, y no cuando X realiza el acto de comprometerse? Creemos que la respuesta radica en la confusión entre una postura causal del compromiso y una normativa. Desde el punto de vista normativo, tanto b) como c) suponen una violación de la norma general "cumple tus compromisos", pero por motivos diferentes. En el caso b), se trataría más bien de la violación de la norma "cuando te comprometas a algo, no debes reconsiderar la meta hasta conseguirla o percibirla imposible". Esta norma también aplica a los compromisos individuales, y está relacionada con la debilidad de la voluntad, el autocontrol y el hecho de que el propio compromiso sirva como motivo para la acción. El caso c), por el contrario, violaría la norma "sé veraz", y se aplica a cualquier acto de comunicación, no sólo a los que implican un compromiso. Las razones para sancionar a X en b) y c) son diferentes: en el primer caso se trata de una traición; en el segundo, de un engaño.

Por otra parte, el no considerar a2) y d2) compromisos satisfechos se sitúa en el enfoque causal, no en el normativo, ya que apela a los motivos de X para realizar A (a pesar de tener una dimensión normativa, como la distinción entre actuar de acuerdo a la regla, o aceptando la regla).

Conclusión

Ambos enfoques, por sí solos, no pueden dar cuenta del compromiso social. Si sólo apelamos al aspecto causal, no podemos explicar las consecuencias normativas de violar el compromiso. Si, por el contrario, sólo atendemos a los aspectos normativos (es decir, al compromiso como generador de deberes y derechos), no es posible dar cuenta de por qué las intenciones de los sujetos son relevantes a la hora de considerar violado o satisfecho un compromiso, ni de las

diferencias entre los compromisos y otros actos comunicativos generadores de derechos y deberes.

La combinación entre ambos enfoques que se suele emplear a la hora de analizar promesas y amenazas es considerar el aspecto normativo como indicador de la existencia de un compromiso, y el aspecto causal para decidir sobre su satisfacción o violación: creemos que esta combinación proviene de la confusión entre los aspectos normativos de cualquier acto de habla, y los específicos del compromiso.

Una combinación adecuada de ambos enfoques debería, por lo tanto, recoger las motivaciones tanto para la el acto de comprometerse como para realizar la acción a la que se está comprometido, así como los aspectos normativos, desprendidos de los derechos y deberes que el compromiso genera.

El resultado de esta combinación, de manera que se respete la continuidad conceptual entre el compromiso individual y el social, sería aceptar a) y b), y probablemente d1), como compromisos, mientras que c) y d2) serían una declaración (falsa) de intenciones. Confundir ambos supone no distinguir entre estafa y error, lo cual, a nivel metodológico, no resulta adecuado.

Referencias bibliográficas

Austin, J.L. (1962), *How to Do Things With Words*, Oxford, Oxford University Press.

Bratman, M. (1999), *Faces of Intention: Selected Essays on Intention and Agency*, Cambridge (MA), Cambridge University Press.

— (2004), 'Three Forms of Agential Commitment: Reply to Cullity and Gerrans' *Proceedings of the Aristotelian Society* 104 (3), pp. 327-35.

Castelfranchi, C. (1996), 'Commitment: from Intentions to Groups and Organizations', *ICMAS-96*, Cambridge (MA), AAAI/MIT Press.

— (2003), 'Grounding We-intentions in Individual Social Attitudes', en Sintonen M., Miller K. (ed.), *Realism in action - Essays in the Philosophy of Social Sciences*, Dordrecht, Kluwer, pp. 195-212.

Cohen, P. R. y Levesque, H. J. (1990), 'Intention is Choice with Commitment', *Artificial Intelligence* 42 (2-3), pp. 213-61.

Elster, J. (2000), *Ulysses Unbound: Studies in Rationality, Precommitment and Constraints*, Cambridge, Cambridge Univ. Press.

Millar, A. (2004), *Understanding People: Normativity and Rationalizing Explanation*, Oxford, Oxford University Press.

Schelling, T. C. (1960), *The Strategy of Conflict*, Cambridge (MA), Harvard University Press.

— (2001), 'Commitment: Deliberate versus Involuntary', en Nesse, R. (ed.), *Evolution and the Capacity for Commitment*, New York, Russell Sage.

Searle, J. (1969), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge, Cambridge University Press.

— (2001), *Rationality in Action*, Cambridge (MA), MIT Press.

Sen, A. (1977), 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory', *Philosophy and Public Affairs* 6 (4), pp. 317-44.
— (1985), 'Goals, Commitment, and Identity' *Journal of Law, Economics and Organization* 1 (2), pp. 341-55.
— (2005), 'Why Exactly is Commitment Important for Rationality?' *Economics and Philosophy* 21 (1), pp. 5-14.

Consideraciones sobre la semántica de Locke*

Luis Fernández Moreno
Universidad Complutense de Madrid
luis.fernandez@filos.ucm.es

La teoría semántica de los términos de género natural predominante en la actualidad es la teoría histórico-causal formulada por Kripke y Putnam en la década de los setenta del siglo pasado. En (1690) Locke propuso una teoría semántica de dichas expresiones – en su terminología, de los términos de *sustancias naturales*[1]–, que es considerada incompatible con aquélla.

Como es sabido, Putnam hizo hincapié en dos contribuciones involucradas en la determinación de la referencia o extensión de los términos de género natural: *la contribución del entorno* y *la contribución de la sociedad* [Putnam (1975), pp. 271 y 245]. Por una parte, la extensión de un término de género natural viene determinada por propiedades subyacentes de los miembros del género pertenecientes a nuestro mundo. Por otra parte, la elucidación de dichas propiedades es objeto de la investigación científica y quienes la llevan a cabo o emplean sus resultados – los expertos – tendrán un mejor conocimiento que el hablante medio acerca de las condiciones de pertenencia a la extensión de un término de género natural. Hay a este respecto, según Putnam, una división del trabajo lingüístico, de acuerdo con la cual el hablante medio está dispuesto a deferir en los expertos la determinación de la extensión de los términos de género natural.

Tras presentar algunos de los componentes fundamentales de la teoría de los términos de sustancia propuesta por Locke, el objetivo de este escrito es examinar si la teoría de Locke puede incorporar la división del trabajo lingüístico y, por tanto, "la contribución de la sociedad", si bien centraré mis consideraciones en un tipo de términos de sustancia que hoy en día suelen ser denominados "términos de sustancias químicas", como "oro" y "agua".

En la teoría semántica de Locke cabe distinguir una teoría del significado y una teoría de la referencia, siendo esta última dependiente de la primera. Las consideraciones de Locke sobre los términos se centran en los términos generales, que incluyen los términos de sustancia. Locke afirma que, exceptuando ciertas expresiones, como las conjunciones y las proposiciones, "las palabras en su significación primaria o inmediata nada significan, salvo las ideas que están en la mente de quien las usa" (3.2.2). Ahora bien, las palabras pueden significar de manera mediata otras entidades. Por una parte, significan – o se supone que significan – las mismas ideas en las mentes de los oyentes. Por otra parte, en la medida en que las ideas representan entidades no-mentales, las palabras, de

* La elaboración de este escrito ha contado con la financiación otorgada por el Ministerio de Ciencia e Innovación al proyecto FFI2008-03092.
[1] Al ocuparme de la teoría de Locke entenderé por "sustancia" las sustancias naturales.

manera mediata, pueden referirse a estas entidades. De este modo los términos generales significan ideas generales, y así concibe Locke los géneros (*kinds*) o tipos (*sorts*), pero se refieren a las entidades que concuerdan con tales géneros.

El tipo más básico de ideas son las *ideas simples*, cuyas fuentes son la sensación y la reflexión. Así mediante los sentidos obtenemos ideas de cualidades sensibles, como formas, colores, sabores, etc. y mediante la reflexión obtenemos ideas de las operaciones de nuestras mentes, como dudar, creer, etc. Es importante distinguir entre ideas y cualidades, y en relación a estas últimas entre cualidades *primarias* y *secundarias*. Las primeras – como solidez, extensión, figura, número, movimiento y reposo – están en los objetos mismos y en sus partículas componentes, mientras que no ocurre así con las segundas – como colores, sonidos, sabores, etc. –, que son simplemente *potencias* de los objetos para producir en nosotros ciertas ideas de sensación por medio de sus cualidades primarias y, en última instancia, de las cualidades primarias de las partículas que componen tales objetos. Locke, siguiendo a Boyle, acepta una concepción corpuscular de la materia según la cual ésta consta de partículas no-sensibles o corpúsculos.

Sobre la base de las ideas simples, la mente lleva a cabo diversos tipos de acciones para construir otras ideas. Dos de estas acciones son la composición y la abstracción de ideas. Por medio de la primera obtenemos ideas *complejas*, construidas mediante la combinación de ideas simples. Mediante la abstracción obtenemos ideas generales o abstractas. Las *ideas de substancias* son ideas complejas abstractas y se diferencian de otras ideas complejas porque se supone que representan entidades que existen con independencia de la mente. No obstante, la teoría de Locke acerca de las sustancias es fundamentalmente, y así lo asumiremos en lo siguiente, una teoría de los *géneros* o *tipos* de sustancias.

Ahora bien, puesto que las ideas simples que componen las ideas de sustancia son experimentadas como proviniendo conjuntamente y coexistiendo, suponemos que las cualidades o propiedades que producen esas ideas dependen de la *esencia real* de los objetos – la constitución (corpuscular) interna compartida por los objetos pertenecientes al mismo tipo –. A este respecto Locke establece una contraposición entre la esencia real y la *esencia nominal* de las sustancias. La esencia nominal de una sustancia es la idea compleja abstracta que tenemos de la sustancia, idea que conocemos, mientras que según Locke no conocemos la esencia real de una sustancia y no tenemos una idea de ella, aunque la esencia real sea el fundamento de las propiedades cuyas ideas constituyen la esencia nominal.

El significado de un término de sustancia es la esencia nominal correspondiente, y puesto que ésta es una idea compleja, será analizable en base a sus ideas simples componentes. Así el significado del término "oro" es la idea compleja abstracta o esencia nominal del oro, que está compuesta de ideas tales como las de ser un cuerpo amarillo, de cierto peso, maleable, fungible, etc., y un objeto pertenece a la extensión del término "oro" si y sólo si tiene las propiedades cuyas ideas constituyen la esencia nominal del oro. Formulado de manera general, la referencia de un término de sustancia viene determinada por la esencia nominal.

No obstante, Locke hace hincapié en que los términos son signos *arbitrarios* de ideas, por lo que es posible que las ideas significadas por el mismo término en su uso por distintos hablantes sean diferentes. Locke reconoce no sólo que esto es posible, sino que de hecho ocurre así; por ejemplo, concede que distintos hablantes asocian o pueden asociar con el término "oro" una esencia nominal diferente (vid., p.ej., 3.6.35 y 3.9.17). Sin embargo, hay una *tensión* dentro de la posición de Locke, pues él afirma que el principal uso del lenguaje consiste en la *comunicación* de ideas, pero no puede haber comunicación si las mismas palabras significan ideas diferentes cuando son usadas por el hablante y el oyente. El problema que se suscita es el de cómo aliviar la tensión entre el vínculo arbitrario entre términos e ideas y el uso principal del lenguaje consistente en la comunicación de ideas. Un indicio de la posible respuesta de Locke a este problema se encuentra en el siguiente pasaje:

> "Los hombres aprenden nombres, y los emplean en conversación con otros hombres, sólo para ser entendidos, lo cual únicamente se logra cuando, *por costumbre o consenso*, el sonido que produzco por medio de los órganos del habla provoca en la mente de quien lo escucha la idea a la cual lo aplico en la mía cuando lo profiero." (3.3.3; mi cursiva).

En un sentido similar Locke apela a menudo al *uso común* de las palabras (vid., p.ej., 3.2.8, 3.4.11, 3.6.51 y 3.11.11), que vincularía las mismas palabras en el uso por parte de distintos hablantes con las mismas ideas. Pero si esto es así, Locke está presuponiendo que en ese "uso común" el vínculo entre palabras e ideas deja de ser arbitrario y pasa a ser *convencional*. De este modo Locke asume que, al menos por regla general, en el discurso cotidiano los hablantes involucrados están empleando las palabras en su "uso común" y que las ideas significadas por tales palabras son compartidas.

No obstante, la apelación al uso común no siempre resuelve el problema mencionado:

> "[A] veces […] resulta necesario, para fijar la significación de las palabras, declarar cuál es su significado, ya sea cuando el uso común lo ha dejado en la incertidumbre o en la vaguedad (como sucede con la mayoría de los nombres de ideas muy complejas), ya sea cuando un hombre las emplea en un sentido un tanto peculiar a él mismo, ya sea cuando el término, siendo decisivo en el discurso y aquel sobre el cual principalmente gira, está expuesto a duda o a equívocos." (3.11.12).

Locke reconoce que, con objeto de evitar malentendidos, el hablante puede explicar el significado de las palabras que usa y, dependiendo del tipo de palabras, puede recurrir a distintos procedimientos. En el caso de los términos que significan ideas simples, puede recurrir al empleo de términos sinónimos, o de descripciones de la cualidad que produce en la mente la idea en cuestión o a la ostensión de dicha cualidad (3.11.14). En el caso de los términos de sustancia, Locke propone recurrir a la ostensión y a la definición (3.11.19 ss.).

No obstante, especialmente por lo que concierne a la determinación del significado y, por tanto, de la referencia de los términos de sustancia, un papel destacado les estaría encomendado a aquellos miembros de nuestra comunidad lingüística más versados en su significado que el hablante medio. Locke afirma:

> "Sería de desear […] que los hombres versados en las investigaciones físicas, y conocedores de los diversos tipos de […] [sustancias], registraran las ideas simples en las cuales observan que los individuos de cada tipo constantemente concuerdan. Esto remediaría, en mucho, esa confusión que se origina en la circunstancia de que diversas personas aplican el mismo nombre a una colección de un número menor o mayor de cualidades sensibles, según estén mejor o peor familiarizados con las cualidades de cualquier tipo de cosas que caen bajo una misma denominación, o según hayan tenido mayor o menor esmero en examinar dichas cualidades." (3.11.25).

Ahora bien, puesto que ésta es una empresa *deseable*, es una que Locke considera irrealizada y, más aún, irrealizable en su época, lo cual no es óbice para que admita que "el herrero o el joyero conocen por regla general ["las verdaderas ideas complejas de […] sustancias [como el oro o el diamante] mucho mejor que el filósofo [y, cabría añadir, mucho mejor que el hablante medio – LFM]." (2.23.3).

De este modo Locke admite que hay miembros de nuestra comunidad lingüística que son mejores conocedores que el hablante medio del significado y, por tanto, de la referencia de los términos de sustancia, y puesto que esto habría de ser concedido igualmente por el hablante medio, cabe asumir que en caso de duda éste estaría dispuesto a deferir en ellos.

Ahora bien, hay una diferencia entre Locke y Putnam motivada, al menos en parte, por las diferentes épocas en las que vivieron. Según Putnam sólo con el surgimiento de la ciencia los términos de género natural pasaron a estar sujetos a la división del trabajo lingüístico, y en el caso concreto del término "agua" esto ocurrió con el surgimiento de la química [1975, p. 228]. Por su parte, Locke era pesimista acerca del alcance del conocimiento científico sobre las sustancias naturales (4.3.26). Sin embargo, incluso antes del surgimiento de la ciencia cabría distinguir entre expertos y no-expertos – en un sentido amplio – acerca de la referencia de los términos de sustancia. En este sentido Locke acepta que hay miembros de nuestra comunidad lingüística que tienen mejores ideas de sustancias que otros (2.23.7) y, como en el caso mencionado del herrero y del joyero, cabe asumir que el hablante medio estaría dispuesto a deferir en el juicio de ellos.

De este modo, dejando de lado el pesimismo de Locke acerca del alcance de la ciencia, no hay razón por la que la teoría de Locke no pudiese incorporar la tesis de la división del trabajo lingüístico y, por tanto, la contribución de la sociedad a la determinación de la referencia de los términos de sustancia.

Referencias bibliográficas

Locke, J. (1690), *An Essay concerning Human Understanding*, ed. P. H. Nidditch, Oxford, Clarendon Press, 1975. (Citado por número de libro, de capítulo y de sección).

Putnam, H. (1975), 'The meaning of 'meaning'', en H. Putnam, *Mind, Language and Reality*, Cambridge, Cambridge University Press, pp. 215-71.

!"#$%&$%"'()($*(+,-.*$/'(0$(*-%(12'*'([3]

!

!"#$%&'()*)+',%-($.%
"#$%&'($)*)!+,)-#./*!0&!1*'2&3.#*!
.34*56&'#*#0&78,*952*)!

!

:&! *2,&'0.! 2.#! 3*(!)&.';*(! '&<'&(&#)*2$.#*3&(! 0&!!3*&'2&<2$=#>! &3! 2*'?2)&'! 6&#./@#$2.! 0&! ,#*! &A<&'$&#2$*! 2.#($()&! B.! /&'*/&#)(<&'%$&#&B! &#! &3! 2.#)&#$0.!'&<'&(&#)*2$.#*3!0&!3*!&A<&'$&#2$*!5'3$(2*!D,&!&A<&'$&#2$*(!2.#!&3! /$(/.!2.#)&#$0.!0&9&#!)&#&'!&3!/$(/.!2*'?2)&'!6&#./@#$2.5!+0&/?(>!0&!*2,&'0.! 2.#! &3! '&<'&(&#)*2$.#$(/.>! 3*(! &A<&'$&#2$*(! (.#!3&)*/&#)&! 0$?6*#*(! &#! &3! (&#)$0.! 0&! D,&! 2,*#0.! (&! *)$&#0&! $#)'.(<&2)$%*/&#)&*(>! 3.! D,&! (&! &#2,&#)'*! #.!(.#!/?(!D,&!3*(!<.<$&0*0&(!<&'2&<)$93&(!0&!.9E&)'$&*3!<&'2$9$0.5!

F*!/*G.';*!0&!3*(!.9E&2$.#&(!*!&()&!$<.!0&!)&0&!3*!<&'2&<2$=#!(&!23*($6$2*#! &#! 0.(! 2*)&4.';*(H! 2*(.(! 0&! $#%&'($=#! 0&! 1$>! &#! 3.(! 2,*3&(! (&! (,<.#&! D,&! 3.(! 2.#)&#$0.(! 6&#./@#$2.(! 0&! 0.(! &()*0.(! (&! &#2,&#)'$#%&')$0.(! &#! '&3*2$=#! *! (,! 2.#)&#$0.!'&<'&(&#)*2$.#*3!I2./.!&E&/<3.!3*!$0&#)$0*0!(&<&2)'.!$#%&')$0.!0&!JG0#&G! JK.&/*L&'>!MNOM>!G!&3!2*(.!*#?3.4.!0&!3*!P$&'!$Q#0&!R&0!13.2L>!MNNSTU!G! 2*(.(!0&!/0$"1$*,(&#)&(>!&#!3.(!2,*3&(!(&!(,<.#&!D,&!0.(!&()*0.(!&#&#!&3!/$(/.! 2.#)&#$0.!'&<'&(&#)*2$.#*3!*,#D,&!,#.!0&!&33.(!&()&!&2.#)&#$0.!6&#./@#$2.!&#! *9(.3,).!I<.'!&E&/<3.>!3.(!2*(.(!,2341'5T5!

V.'!3.!D,&!'&(<&2)*!*!3.(!2*(.(!0&!&(<&2)'.!$#%&')$0.(!$#%*'$*(!3*(!'@<3$2*(!D,&! (&!K*#!.6'&2$0.5!F*!<.($9$3$0*0!0&3!&(<&2)'.!$#%&')$0.!&#0&!/.()'*'!D,&>!2.#)'*! 3.! D,&! (&! 0&6$&#0&! 0&(0&! &3! '&<'&(&#)*2$.#$(/.! &#! 2*/(*! (! D,&!)$&#&#! 2.#)&#$0.(! $0@#)$2.(! <,&0&#! ($#! &/9*'4.! 0$6&'$'! &#! 2*'?2)&'! 6&#./@#$2.5! J,<=#4*(&! D,&! &3! &(<&2)'.! 0&! 2.3.'! 0&! +! &()?! $#%&')$0.! '&(<&2).! *3! 0&! 15! W,*#0.!+!/$'*!.9E&).(!'.E.(!)$&#&!3*!&A<&'$&#2$*!D,&!#!1!2,*#0.!%&!.9E&).(! *7,3&(5!F*(!&A<&'$&#2$*(!0&!+!G!0&!1!3&(!<&'2$9$0&*'!3.(!/$(/.(!.9E&).(! <.'!&3!2.3.'>!<&'.!3.!D,&!&()&!&A<&'$/&#).!/&#)*!3$#0,2&!*!*2&<)*'!&(!D,&!&3! 2*'?2)&'!6&#./@#$2.!0&!(,(!&A<&'$&#2$*(!I!&(!0$#().!D,$&'&!0&2$'!D,&!K*G! *34.!/?(!&#!&3!2*'?2)&'!6&#./@#$2.!0&!,#*!&A<&'$&#2$*!D,&!&3!2*'?2)&'!&#)*5!F*(! &A<&'$&#2$*(!)$&#&#!*!0&!(,!$"&0,2$93&(5!!

+!&()*!0&6&#(*!0&!3*!/0$"1$(&!K*!'&<3$2*0.!D,&!/&'*/&#)&!*<&3*!*!,#*!0,0.(*! $#),$2$=#5!XV.'!D,@!*34,$&#!0&9&';*!(&#)$'!&3!/$(/.!*'?2)&'!D,&!&(!<.($93&!D,&! 3*(! &A<&'$&#2$*(! 0&! 2.3.'! 0&! +! (&! &#2,&#)'&#! $#%&')$0*(! '&(<&2).! *! 3*(! 0&! 1! /?(!D,&!<.'D,&!(./.(!2*<*$&'$!!(&'!+!.!1!G!)&#&'!3*(!&A<&'$&#2$*(!0&! *7,3!/$(/.!'*/.(!*34,!'.E.Y!R$!&3!K&2K.!0&!D,&!<,&0*!)&#&'(&!'&(<&2).!*!&3! K&2K.! 0&! D,&! <,&0*! 2.#2&9$'(&! &()*93&2&! (,! <.($9$3$0*0)'*3/&#)&>! *! &()*(!

!!!

[3]! C()&!)'*9*E.!(&!K*!9&#&6$2$*0.!0&!3*!6$#*#2$*2$=#!0&3!0&!W$&#2$*!&!Q##.%*2$=#!*!)'*%@(!0&3!<'.G&2).!0&!$#%&()$4*2$=#!\\Q]SSO^S_M_`^WS]^S]!G!0&!3*!*G,0*!0&!3*!1&#&'*3!W*)*3,#G*!*3!a',<!0bQ#%&()$4*2$=!&#!C<$()&/.3.4$*!$!W$&#2$&(!0&!3*!*W!/&'!$!&(2!a'WadCWWT!Jad]SSN^ Me]O5!

'@<3$2*(! 3&! K*#! (&4,$0.! (,9($4,$&#)&(! 0&6&#(*(&(($&(0&! &3! ?/9$).! 0&! 3*! 0$(2,($=#!/&)*6;($2.^2.#2&<),*35!

f)'.(>!*,#D,&!*2&<)*#!3*!6,&'7*!0&3!K&2K.!0&!D,&2(&2&9$93&>!2,&()$.#*#! D,&!&3!2*(.!0&(2'$).!'&*3/&#)&!3.!(&*5!F*!$#%&0($3#&(<&2)'.!(.3*/&#)&!(&';*! <.($93&!($!&3!&(<*2$.!0&!2.3.'!6,&'*!($/@)'$2.&(&(D,&!(&!$#%$')$&'*!&3!&(<&2)'.! ($#!D,&!*6&2)*(&!*!3*(!K*9$3$0*0&(!0$(2'$/$#*)&3*(<&'2&<).'5!J$#!&/9*'4.>! 2.#)$#,*!&3!*'4,/&#).>!&3!&(<*2$.!0&!2.3.'!#.!<,&(&'!($/@)'$2.5!P.0.(!3.(! 2.3.'&(!)$&#&#>!<.'!&E&/<3.>!(*),'*2$=#!G!9'$&3&(!/.0.!3.(!*/*'$33.(!(.#>! 2./.!,#!).0.>!/?(!9'$33*#)&(!D,&!3.(!*7,3&(!<&'.!<.0&/.(!0$(2'$/$#*'!/?(!*7,3&(! 2.#!0$()$#)*(!(*),'*2$.#&(5!C().!D,&'';*!0&2$'!D(&!(&!$#%$')$&'*#!3.(!&(<*2$.(!0&! 2.3.'!0&!+!2.#!'&(<&2).!1>!&().!*6&2)*';*!*!(,!K*9$3$0*0!0$(2'$/$#*0.'*5!

C3! .9E&)$%.! 0&! /$! 2./,#$2*2$=#! &(! 0&6&#0&'! D,&#&2$(! '&*3&(! D,&! *<,#)*#!&#!3*!0$'&22$=#!0&!3.(!2*(.(!0&!&(<&2)&$$0.5!C3!K&2K.!0&!D,&!#.(! &#6'&#)&/.(!*!2*(.(!2.#!'&*3$0*0!<($2.3=4$2*!3&()*!&3!9&#&6$2$.!0&!#.!)&#&'! D,&!(,()&#)*'(&!&#!$#),$2$.#&(!.!&#!3*!6,&'7*!3.'$*!0&!3.!2.#2&9$93&5!C()&!)$<.! 0&!2*(.(!(&!K*33*#!&#!&#!3!6&#=/&#.!0&!3*!($#&()&(6#50&6$&#0.>!3*!/&E.'! &A<3$2*2$=#!0&!&()&!6&#=/&#.!2./.!6&#=/&#.!<&'2&<)2&($#&3*!*2&<)*2$=#!0&! D,&!0.(!&()*0.(!2.#!&3!/$(/.!2.#)&#$0.!'&<'&(&#)*2$*3!2./<*')&#!,#*!<*')&>! <&'.!#.!).0.>!(,!2.#)&#$0.!6&#./@#$2.5!

F.(!&A<&'$/&#).(!/&#)*3&(!0&3!)$<.!h&(<&2)'.!$#%&')$0.!G!#.!'&)&#0&#!/.()'*!D,&! 3*(!&A<&'$&#2$*(!<&'2&<),*3&(!)$&#&!'*(4.(!#2$#)3&#(!G!D,&!#.(!<&'2*)*/.(! $#)'.(<&2)$%*/&#)&!0&!&().(!'*(4.(5!C().(!&A<&'$/&#&#)*3&(>!0&!K&2K.>!#.(! <'&)&#0&#!2.#%&#2&'!0&!&()*!&A$()&#2$*!*3!/(()2=#.!<.0';*/.(!33&4*'!*! <&'2*)*'#.(!0&!&().(!'*(4.(5!V,&(!9$&#>!3.(!2*(.(!0&($#&()&($*!#.(!/,&()'*#>!&#! <'$/&'!3,4*'>!2./.!2.#!'&4,3*'$0*0!G!($()&/*)$2$0*0&(!<.($93&!<&'2*)*'(&!0&!&(.(! '*(4.(5!C#)$&#0.!D,&!&()&!6&#=/&#.!(.2*%*!3*(!%&#&(!&A)&'#$()*(!0&3! '&<'&(&#)*2$.#$(/.!*3*!%&7!D,&!/,&()'*!D,&!&(!&#&*!3*!*2&<)*2$=#!2./.!<'&/$(*! 0&!3*!)&($(!0&!0$*6*#$0*0!<.'!<*')&!0&!3.(!0&#&3!'&<'&(&#)*2$.#$(/.5!

W./g#/&#)&!3*!<*3*9'*!i($#&()&($*j!0&($4#*!,#*!2.#$2=#!0&!*34,#.(!,E&).(! (&4g#!3*!2,*3!3*!&()$/,3*2$=#!&#!,#*!/.0*3$0*0!(&#3!0*!3,4*'!*!&A<&'$&#2$*(! &#! ,#*! /.0*3$0*0! 0$()$#)*5! J$#! &/9*'4.>!)*/9$@#!2&#($0&'*! ($#&()&($*! *! 3*! 2.#0$2$=#! (&4g#! 3*! 2,*3! 3*! &()$/,3*2$=#! &#! ,#*!3$0*0! (&#(.'$*3! 0*! 3,4*'! *! 0$()$#)*(! 2,*3$0*0&(! 0&! &(*! /$(/*! /.0*3$0*05! V.'!&E3.>! 3.(! 2*(.(! &#! D,&! 3*! %$($=#!0&!3&)'*(!&%.2*!2.3.'&(5!F.(!(,E&).(!($#!0$&#&#!&#).#2&(!&A<&'$&#2$*(! *);<$2*(5!J$#!&/9*'4.>!G!&(.!K*2&!*g#!/?(!$#)&'&#&().(!2*(.(*!&A)'*k&7*! #.(! <*'&2&! 9*()*#)&! 6?2$3! $/*4$#*'! 3*!($#&! &A<&'$&#)*'3*! 0$'&2)*/&#)&!<.'D,&!$#%.3,2'*!2.#E,#).(!0&!(&#(!&6*/3*'&(5!V.'!&E&/<3.>! 3*!($#&()&($*!/?(!2./g#!&(!*D,@33*!&#!D,&!3.(!2*'2(!*36*#,/@'$2.(!(&!<&'2$9&#! 2.#!2.3.'&(5!C()*!2.#0$2$=#!0&!3.(!($#&()&)*(!#(3&>!($#!&/9*'4.>!#$#4,#*! $#)&'6&'&#2$*! &#! (,! 0;*! *! 0;*! I<.'! &33.>! &#! /,2K(!(.(>! #.! (*9&#! 0&! (,! 2.#0$2$=#!&(<&2$*3T5!C().(!(E&).(!0$()$#4,&#/&#)'&!3.(!2.3.'&(!0&!3*(! 3&)'*(! .! #g/&'.(! I&A)&'#.(! *! &33.(! 0$4?/.(3.! *(;T! G3.(! 2.3.'&(! D,&! %$&#&#! ($()&/?)$2*/&#)&!2.''&3*2$.#*0.(!2.#!3*(!3&)'*(!.(#g/&'.(<&'2$9$0.(>!3.!D,&! 3.(!*(&/&E*!*!,#!2*(.!0&!&(<&2)'.!$#%&')$0.!G!#.!'*/&#)&!*!3.!D,&!<.0';*#!(&'!

$9.'(13$#'55! :&! &()&! /.0.>! <.0';*/.(! 0&2$'! D,&! 3*! ($#&()&($*,!&0&! 0&(2'$9$'(&! 2./.! ,#*! 2.#0$2$=#! D,&! <&'/$)&! D,&! 3.(! (,E&).(!)&##4! &A<&'$&#2$*(! 2.#! '&63&E.! '&<'&(&#)*2$.#*3! G! &A<&'$&#2$*(! (&#(.'$*3&(! ($#@$35$0,.(! +! I#.! ($#&()&)*T! G! 1! I($#&()&)*T! &#6'&#)*0.(! *! 3.(! /$(/.(! &();/,3.(! (&! &#2,&#)'*#! &#! 0.(! &()*0.(! $4,*3&(! &#!).0.(! 3.(! *(<&2).(! '&3&%*#)&(! <*'*! ,#*! &!'),*! '&<'&(&#)*2$.#$()*! 0&3! 2.#)&#$0.! G! ($#! &/9*'4.! (=3.! 2.$#2$0&#! <*'2$*3/&#)&! &(,! 6&#./&#.3.4;*5! C(! 2$&').! D,&>!)*3! G! 2./.! *6$'/*#! 3.(! '&<'&(&#)*2$.#$()*(! &(*! 6&#./&#.3.4;*! &()?! 2*'*2)&'$7*0*!<.'!<'.<$&0*0&(!0&3!.9E&).!/*)&'$*35!(=3.!&#!<*')&5!C#!&3!2*(.!0&! 1! K*G! .)'*! <*')&! D,&! (&! *k*0&! ($()&/?)$2*/&#)&5! F*! <$&0*0! .9E&)$%*! D,&! &(! '&(<.#(*93&! 0&! 3*(! &A<&'$&#2$*(! &#! +! G! 1! &(! 3*5! /$(/D,&! &(&! &(! &3! 2*(.>! &3! '&<'&(&#)*2$.#$()*! #.! <,&0&! &A<3$2*'! 3*! 0$6&'&#2$*! &#)'&! &A<&'$&#2$*(! 6&#./@#$2*(!*k*0$0*(!D,&!*0&/?(!(.#!23*'*/&#)&!0$()$%*(!<.'D,&!#.!2./<*')&#! 2.#!.)'*(!&A<&'$&#2$*(!3*!2*'*2)&';()$2*!0&!3*!#($$*0*05!

C#! &3! 0&9*)&! 2.#)&/<.'?#&.! 6,&! a$39&')! m*'/*#! D,$&#)'.0,E.! 3*(! 2.#($0&'*2$.#&(! (.9'&! 3*! 0$*6*#$0*0! .! 3*!)*/9$@#! 0*33)'*#<*'&#2$*! 0&! 3*! &A<&'$&#2$*5!C#!,#!6*/.(.!<*(*E&!*6$'/*H!

> W,*#0.! C3.$(*! %&! ,#! ?'9.3! *#)&! &33*>! 3.(! 2.3.'&(! D,&! &A<&'$/&#)*! (.#!).0.(! &A<&'$/&#)*0.(! 2./.! '*(4.(! 0&3! ?'9.3! G! (,! &#).'#.5! R$#4,#.! 0&! &33.(! (&! &A<&'$/&#)*!2./.!'*(4.(!$#)';#(&2.(!0&!(,!&A<&'$&#2$*5!P*/<.2.!&A<&'$/&#)*! #$#4g#!'*(4.!0&!*34.!2./.!,#!'*(4.!$#)';#(&2.!0&!(, !&A<&'$&#2$*5!C().!)*/9$@#! &(!%&'0*0!0&!,()&0!/$(/.5!I555T![$'&!,#!?'9.3!G!)*!0&!&#6.2*'!(,!*)&#2$=#!&#! '*(4.(! $#)';#(&2.(! 0&! (,! &A<&'$&#2$*! %$(,*35! V'D0&! 4()&0! &#2.#)'*'?! D,&! 3.(! g#$2.(! '*(4.(! *! 3.(! D,&! <,&0*! *)&#0&'! (&'?#! '*(4.(! 0&3! ?'9.3! D,&! (&! <'&(&#)*555!nm*'/*#>!MNNSo5!

W.#! 2*($! 3*! g#$2*! &A2&<2$=#! 0&! FG2*#! I]SS`T>! <3*(&#)&(! 0&3! '&<'&(&#)*2$.#$(/.! 6'&2,&#)&/&#)&! /.)$%*#! (,! <,#).! 0&! %$()*! &#! &()&!)$<.! 0&! 2.#($0&'*2$.#&(! 6&#./&#.3=4$2*(5! +6$'/*#! D,&! ($/<3&/&#)&! #.! (.#! 2*<*2&(! 0&! &#2.#)'*'! &#! (,! &A<&'$&#2$*! #$#4g#! '*(4.! 0&3! 2,&'<.! (2*)&#! /?(! *33?! 0&! 3.(! D,&! 3.(!D,&!(&!'&<'&(&#)*!D,&!)$&#&#!3.(!.9E&).(5!!

"#*!.<.(2=#!&()?#0*'!*!3*!)&($(!0&!3*!0$*6*#$0*0!&()?!D,&!.6'&2$=!R&0!13.2L5! V.'! &E&/<3.>! @()&! *<&3*! *! &A<&'$&#2$*(! %$(,*3&(! (&#(*2$.#&(! 0&! 2.3.'! 2'&*0*(! 2,*#0.! (&! <'&($.#*! &3! .E.! 2.#! 3.(! <?'<*0.(! 2&"*0.(T! &#! 3*(! D,&! i3*! 0$*6*#$0*0!0&!3*!<&'2&<2$=#!&(!/,2K.!/&#.(!<'.#,#2$*0*! I13.2L>!MNN_>!<5!peT5! "#*! '&(<,&()*!)*/9$@#! 23?($2*! *! &().(! <'&)&#0$0.(! &E&/<3.(! &(! D,&>! 0&! K&2K.>!&().(!2*(.(!#.!(.#!&A<&'$&#2$*(!%$(,*3&(! (3.&! 0&'&2K.5!J&!(,<.#&!D,&! *,#D,&! K*G*! 2*(.(! &(<&2$*3&(! &#! 3.(! D,&! 3*! &A<&'$&#2$*'2&<),*3! #.! &(!)'*#(<*'&#)&! .! 0$?6*#*>! <.0';*! *g#! 0&6&#0&'(&! D,&! 3*! &A<&'$&#2$*! %$(,*3&(! 2.)$0$*#*(!(;!3.!(.#5!!V,&(!9$&#>!($!@()*!#.!/&2&'3*!/&E.'!'@<3$2*!*!&()&!)$<.!0&! 2.#)'*&E&/<3.(!/,2K.!/&#.(!<,&0&!(&'!*2&<)*93&!2,*#0.!(&!&A*/$#*#!3.(!2*(.(! 0&! ($#&()&($*5! F*(! &A<&'$&#2$*(! %$(,*3&(! ($#&()@($2*(! G! 2.)0*#*(! 2./.!2,*3D,$&'!&A<&'$&#2$*!%$(,*3!0&!,#!(,E&).!$#&()&)*![M]5

[M]! d&2$&#)&/&#)&>! q*2L(.#! I]SS_T! G! [.3G#&,A! I]SSNT! .6'&2&#! *'4,/&#).(! 2.#),#0&#)&(! <*'* /.()'*'!D,&!&3!'&<'&(&#)*2$.#$(/.!#.!0&9&';*!,)$3$7*'(&!<*'*!&/$(*(!0&!2*'?2)&'!6&#./&#.3=4$2.!<*'*!

*(*1)* ***most of f*** *: Rngx ⤖ Rngy s.t.*

<x,f(x)> ∈ Own & ∀v∀w [f(v)=f(w) → (v=w or (∃a∈Rngy : <w,a>∈ Own & ∀z f(z)≠a))]

most of x *(<x,f(x)>∈ Own → <x,f(x)>∈ Beat)*

o lo que es equivalente:

*(*2)* ***most of f*** *: Rngx ⤖ Rngy s.t.*

<x,f(x)>∈ Own & ∀v∀w [f(v)=f(w)→ (v=w or ∀a (<w,a>∈ Own → ∃z: f(z)=a))]

most of x *(<x,f(x)>∈ Own → <x,f(x)>∈ Beat)*

donde *Rngx* simboliza el conjunto de hombres del modelo que tienen burro, *Rngy* el conjunto de burros del modelo que pertenecen a algún hombre, y ***most of x*** toma sus valores en *Rngx*. A cada una de las aplicaciones ***f*** descritas en las dos primeras líneas de *(*1)* y *(*2)* la llamaremos *situación asociada al antecedente del condicional (*)*. Se trata de un tipo de aplicación a la que imponemos el ser 'lo más inyectiva posible', significando esto que solamente podremos conectar dos hombres distintos a un mismo burro si no existe un burro diferente, y no conectado a ningún otro hombre, al que también podamos asociar alguno de estos dos hombres.

Para hacer una lectura apropiada del condicional habremos de tratar por un lado con conjuntos específicos de pares hombre-burro y por el otro con individuos de *Rngx*. Esto es, un cuantificador sobre *situaciones* y otro sobre elementos.

Nuestra propuesta presenta al menos dos ventajas:

1. A diferencia de DRT y GTS, nuestra interpretación da el valor de verdad intuitivo (falso) al enunciado *(*)* cuando éste es proferido en el modelo (M).

En este caso podemos distinguir hasta diez *situaciones* ***f*** diferentes (entre las cuales **f'**) asociadas al antecedente del condicional *(*)*. El grafo de todas estas

situaciones está constituido por diez pares de elementos hombre-burro en los que, en su mayoría (nueve), el hombre no pega al burro. De acuerdo pues con nuestra propuesta, el enunciado habrá de ser *falso*.

2. Nuestro tratamiento permite distinguir entre voz activa y voz pasiva:

(*) Usually, if a man owns a donkey, he beats it.

(') Usually, if a donkey is owned by a man, it is beaten by him.*

Nuestra interpretación de *(*')* es análoga a la que hacíamos de *(*)*:

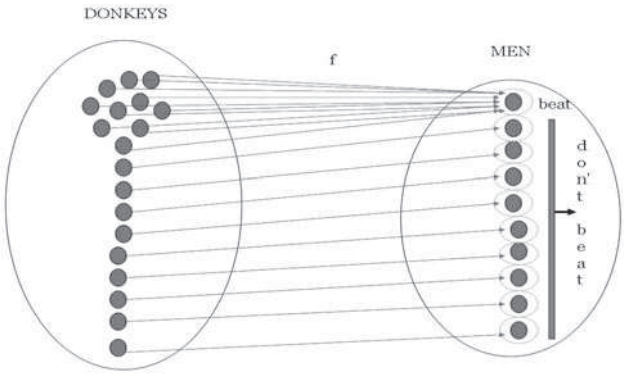
('1)* ***most of f*** *: Rngx → Rngy s.t.*

$<x,f(x)> \in Owned \ \& \ \forall v \forall w \ [f(v)=f(w) \rightarrow (v=w \text{ or } (\exists a \in Rngy : <w,a> \in Owned \ \& \ \forall z \ f(z) \neq a))]$

most of x $(<x,f(x)> \in Owned \rightarrow <x,f(x)> \in Beaten)$

donde, esta vez, *Rngx* simboliza el conjunto de burros que pertenecen a algún hombre y *Rngy* es el conjunto de hombres del modelo que tienen burro.

El enunciado *(*')* resulta ser intuitivamente verdadero en el mismo modelo en el que *(*)* era intuitivamente falso. De nuevo, nuestra interpretación da cuenta de esta diferencia y consigue dar el valor de verdad adecuado a *(*')*: hay una única *situación* **f** asociada al antecedente del condicional *(*')*. El grafo de esta *situación* consta de diecinueve pares de elementos burro-hombre en los que, en su mayoría (diez), el burro es pegado por el hombre. El enunciado habrá de ser pues verdadero de acuerdo con nuestra interpretación.



Estos son, pues, dos de los aspectos positivos que presenta nuestra propuesta. La cuestión que se plantea a continuación es la siguiente: ¿Acaso no sería posible reformular esta idea sin necesidad de acudir a un concepto de *situación* como el que se ha planteado aquí? ¿Es realmente necesario hablar de *situaciones*? ¿Qué decir por ejemplo de la posibilidad de dar una interpretación como la que sigue, *(P)*, a los enunciados *(*)* y *(*')*?

*(P) (*)**mostofx***[(Man(x)^∃y(Donkey(y)^Own(x,y))), (∀y((Donkey(y)^Own(x,y)) →Beat(x,y)))][1]

Esto es: 'la mayor parte de los hombres que poseen un burro, pegan a todos sus burros.'

(P) (') **mostofx***[(Donkey(x)^∃y(Man(y)^Own(y,x))), (∀y((Man(y)^Own(y,x)) → Beat(y,x)))]

Que puede leerse: 'la mayor parte de los burros que pertenecen a algún hombre, son golpeados por todos sus dueños.'

La propuesta *(P)* proporciona, como la nuestra, el valor de verdad intuitivo a ambos enunciados, *(*)* y *(*')*, en el modelo (M). De hecho, podemos observar con cierta facilidad que si un enunciado *(*)* [o *(*')*] es verdadero según la interpretación *(P)* en un modelo, también lo será de acuerdo con la nuestra. El recíproco, sin embargo, no es cierto. Veámoslo. Supongamos que, por ejemplo, *(*)* fuera proferido en un nuevo modelo (M') sometido a las siguientes condiciones:

- Dos hombres poseen burros.
- Uno de los hombres posee dos burros de entre los que pega tan sólo a uno.
- El hombre que resta es también propietario del único burro al que no pegaba el anterior. Él sí que le pega. No posee más burros.

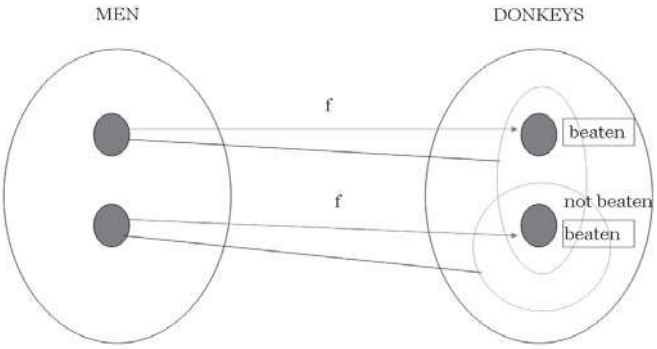
Como ocurría con anterioridad, nuestra propuesta consigue proporcionar el valor de verdad intuitivo (verdadero) al enunciado *(*)* cuando este es proferido en el modelo (M'). Esto se debe a que, dado el modelo (M'), no hay más que una *situación* posible **f** asociada al antecedente de *(*)* (ver gráfico, a continuación). Puesto que todos los hombres (x) del grafo pegan a sus respectivos burros (**f**(x)), el enunciado es verdadero.

[1] Seguimos la notación de [Barwise y Cooper (1981)]:

Most *({ x : Man(x) ^ ∃y (Donkey(y) ^ Own(x,y)) }) ({ x : ∀y (Donkey(y) ^ Own(x,y) → Beat(x,y)) })*, según la cual

‖***Most*** *({ x : A(x) }) ({ x : B(x) })*‖ = true ssi ‖*{ x : B(x) }*‖ ∈ ‖***Most*** *({x : A(x)})*‖, donde

‖***Most*** *({x : A(x)})*‖ = *{ X : | X ∩ { x : A(x) } | > | { x : A(x) } – X | }*



Por el contrario, el mismo enunciado *(*)* proferido en el mismo modelo (M') tendrá valor de verdad 'falso' según la propuesta *(P)*. La razón es la siguiente: tan sólo uno de los hombres que tienen burro les pega a todos (de hecho uno sólo). De ahí que no sea el caso que la mayor parte de los hombres peguen a todos sus burros y, como consecuencia, que tampoco *(*)* sea verdadero de acuerdo con la interpretación *(P)*.

Pareciera pues que la interpretación que hemos dado al condicional se comportara mejor que *(P)*. Pero ¿y si modificáramos en algo *(P)* dando lugar a la siguiente nueva interpretación *(P')* del condicional?

(P')(*) ***mostofx*** [(Man(x)^∃y(Donkey(y)^Own(x,y))), (∃y(Donkey(y)^Own(x,y)^Beat(x,y)))]

Esto es: 'la mayor parte de los hombres que poseen un burro, pegan a alguno de ellos.'

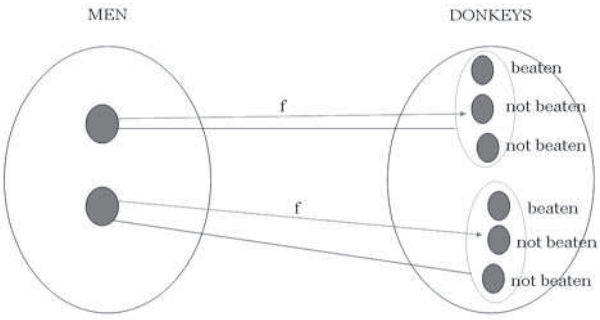
(P')(')* ***mostofx*** [(Donkey(x)^∃y(Man(y)^Own(y,x))), (∃y(Man(y)^Own(y,x)^Beat(y,x)))]

Que puede leerse: 'la mayor parte de los burros que pertenecen a algún hombre, son golpeados por algunos de sus dueños.'

En este caso, *(P')*, es incluso más sencillo de lo que lo fue con *(P)* encontrar un contra-ejemplo. Consideremos, por ejemplo, el modelo (M'') descrito de la siguiente forma:

- Dos hombres poseen burros.
- Cada uno de los hombres posee tres burros diferentes.
- Ambos pegan tan sólo a uno de sus burros.

De nuevo, como ocurría en los casos precedentes, nuestra interpretación del condicional proporciona el valor de verdad intuitivo (falso) del enunciado *(*)* cuando este es proferido en el modelo (M''):



Por el contrario, la interpretación *(P')* asigna a *(*)* el valor de verdad 'verdadero'. La razón es que la mayor parte de los hombres del modelo (M'') (los dos hombres) golpean a alguno de sus burros (de hecho, a uno sólo). Nuestra propuesta parece consolidarse así como la más adecuada a la intuición.

Referencias bibliográficas

Barwise, J. and Cooper, R. (1981), 'Generalized quantifiers and natural language', *Linguistics and Philosophy* 4, pp. 159-219.

King, J. C. (2004), 'Anaphora', *Stanford Enciclopedia of Philosophy*, <http://plato.stanford.edu/entries/anaphora> (23/11/2006).

Sandu, G. (1997), 'On the theory of anaphora: Dynamic Predicate Logic vs. Game-Theoretical Semantics', *Linguistics and Philosophy* 20, pp.147-74.

Van Eijck, J. and Kamp, H. (1997), 'Representing Discourse in Context', in van Benthem, J. and ter Meulen, A. (eds.), *Handbook of Logic and Language*, The MIT Press, Cambridge, Massachusetts, pp. 179-238.

Cognición emocional

Emilio García Buendía
Universidad Complutense de Madrid
egarciabue@yahoo.es

John Searle, en su obra *Mentes, cerebros y ciencia*, formula una crítica que merece la pena retomar. Alega este autor que las explicaciones de la mente de corte funcionalista y que toman al computador como modelo no explican la realidad humana dado que el ente biológico no tiene las mismas características que el ente mecánico denominado computador.

Pero retomando la crítica de Searle cabría preguntarse: los modelos computacionales, de por ejemplo la función racional, ¿explican realmente la función racional humana? Cabría la posibilidad de considerar que se ha explicado qué es razonar pero no haber explicado en qué consiste el razonar humano.

Siguiendo este hilo conductor y siguiendo la metodología interdisciplinaria propia de las ciencias cognitivas, la neurología nos muestra tres síndromes sobre los que merece la pena detenerse. Dichos síndromes son el *síndrome de Capgras* y el *síndrome de Cotard* y los *síndromes de irrealización* y *despersonalización*. Autores como Ramachandran y Blakeslee, defienden en su obra *Phantoms in the brain,* que proporcionan una evidencia clara y empírica en la línea de considerar que la cognición humana se encuentra *teñida* de emoción por lo que cognición y emoción deben ser considerados como un continuo indisoluble.

Planteamiento de la crítica de Searle al simbolismo

Searle [7:35], en su obra *Mentes, cerebros y ciencia* arranca su razonamiento exponiendo lo que él considera el núcleo fuerte del programa simbolista, a saber, que la inteligencia humana queda explicada al considerarla como una mera manipulación de símbolos físicos en sentido estricto y literal de la expresión.

Más adelante, Searle pasa a formular su crítica al programa simbolista presentado como contra-argumento fundamental la idea de que a partir de la sintaxis (mera manipulación de símbolos) jamás puede emerger la semántica [7:37 y 39].

A partir de este momento Searle formula, para sustentar su crítica, el ya famoso experimento mental de la habitación china. Si bien el argumento de que la sintaxis no es suficiente para la semántica constituye el núcleo de la crítica de este autor, en la obra referida apunta de pasada una línea crítica que ha sido poco explorada y que se desea resaltar aquí.

Para dicho programa de investigación, pensar es procesar información y procesar información es solamente trasiego de símbolos. El objeto a explicar es la cognición entendiendo por tal cualquier proceso del pensamiento que comprendería fenómenos tales como el razonamiento, la memoria, la resolución de

problemas, lenguaje, etc. y que excluiría de su ámbito toda el área emocional del sujeto.

La tesis que se plantea aquí es que la cognición, en los entes biológicos, sí que consiste en procesos de información pero que dichos procesos se encuentran todos ellos *teñidos* o mezclados de factores emocionales lo cual forma parte indisoluble de dicha información. Y la razón de ello habría que buscarla en razones puramente evolutivas dado que la información emocional o cognición emocional reporta ventajas adaptativas que permiten a los entes biológicos grandes mejoras en la tarea de la supervivencia. La cognición emocional permite, en este caso al ser humano, una mayor adaptación al entorno y vencer por sobrevivir con grandes ahorros de energía.

A continuación se presentarán tres síndromes neurológicos denominados: a) *síndrome de Capgras*, b) *síndrome de Cotard* y c) *síndromes de irrealización y despersonalización* para comprobar en qué medida permiten fundamentar la tesis propuesta.

El *síndrome de Capgras*

El denominado *síndrome de Capgras* fue identificado por primera vez por el psiquiatra francés Jean Marie Joseph Capgras (1873-1950) [3]. Capgras se encontró con una paciente de 74 años que afirmaba que su marido había sido reemplazado por otra persona si bien reconocía sin dificultad a otros individuos; por eso motivo se le conoce también como el *síndrome del sosias*. Este síndrome se caracteriza básicamente por el hecho de que el paciente no presenta emoción alguna ante la presencia de un familiar directo por lo que dicha carencia afectiva la interpreta racionalmente sosteniendo que dicha persona no es la que parece ser sino otra que se le parece, un *sosias*. La ausencia absoluta de respuesta emotiva se ha podido comprobar debido a la ausencia completa de respuesta galvánica en la piel por lo que las personas estudiadas no habrían simulado su ausencia emocional.

Con posterioridad a su descripción y del posterior estudio de otros pacientes con el mismo síntoma producido por lesiones cerebrales se especuló con la posibilidad de que este síndrome se encontrara asociado con el síndrome de *prosopagnosia* el cual consiste en la incapacidad por parte del afectado de reconocer las caras aunque no tenga problema alguno para reconocer cualquier otro tipo de objetos. A pesar de esta primera hipótesis, Bauer [1] mostró que pacientes con la capacidad para reconocer rostros familiares intacta presentaban el síndrome de Capgras por lo que supuso que deben existir dos vías para el reconocimiento de caras, uno consciente y otro inconsciente.

Posteriormente, Hadyn Ellis y Andy Young [4] plantearon la hipótesis de que el sistema consciente necesario para el reconocimiento de caras estuviese intacto estando dañado únicamente el sistema nervioso responsable del surgimiento de emociones al encontrarse frente a rostros familiares. Recientemente, William Hirstein y V. S. Ramachandran [5] informaron del estudio de un caso en el que un

paciente con el síndrome de Capgras era perfectamente capaz de reconocer todas las caras y sentir emociones si bien dichos sentimientos no surgían ante familiares.

A la vista de toda la evidencia anterior, V. S. Ramachandran [6] ha planteado la siguiente hipótesis explicativa de este extraño síndrome. Si la conexión entre el reconocimiento de lo percibido y las áreas emocionales falla, entonces cualquier persona reconocería la cosa percibida pero no se produciría ningún asociación emocional. En el *síndrome de Capgras* se daría un fallo de este tipo que consistiría en una desconexión entre las vías visuales y emocionales referidas especialmente a familiares. La consciencia, ante esta ausencia emocional que sería esperable al ver a cualquier familiar, realizaría una interpretación *post-hoc* consistente en afirmar que esa persona percibida no es su familiar, lo han sustituido, siendo un doble.

Este síndrome afecta únicamente a la asociación afectiva en el sentido indicado, visón-emoción, pero existe otro síndrome en la misma línea que abarca aspectos más amplios.

El *síndrome de Cotard*

El denominado *síndrome de Cotard* fue descrito por primera vez por el neurólogo francés Jules Cotard (1840-1889) en una lección impartida en 1891 en París al que denominó *delirio de negación* si bien, posteriormente, se le ha designado con el nombre de trastorno *delirante nihilista*. La característica principal de este síndrome es que el paciente tiene la creencia de que está muerto, que no existe, que se está pudriendo o que le faltan sus órganos internos o ha perdido su sangre; raramente pueden incluir delirios de inmortalidad con la creencia de estar condenado a vivir eternamente para sufrir sin cesar. En sus casos más graves estos pacientes niegan la existencia de su propio yo con sentimientos de disolución del mismo. Probablemente, el estudio más amplio sobre este síndrome es el realizado por Berríos [2] .

Un aspecto de este síndrome llamativo es la resistencia extrema que presenta al pensamiento racional de tal modo que estos pacientes son completamente refractarios a cualquier tipo de argumentación.

La explicación más plausible que se maneja, siguiendo a Ramachandran [6] es que, a diferencia de lo que ocurre con el *síndrome de Capgras*, ahora son todas las conexiones de todos los sentidos los que se encontrarían interrumpidos respecto a las áreas emocionales y no sólo el de la visión. De este modo, el mundo que rodea al paciente se encuentra absolutamente desposeído de significado emocional, lo cual necesita ser interpretado por el sujeto de alguna manera, dando lugar a dichas verbalizaciones ya mencionadas en el sentido de afirmar que se está muerto o con las mencionadas disoluciones de su yo. El hecho de que la resistencia a cualquier tipo de racionalización sea tan elevada apuntaría en el sentido de que la emoción prima sobre lo racional, es decir, que en el ser humano lo emocional tendría una prioridad lógico-temporal.

La conclusión por tanto de este síndrome es que cuando las conexiones cognitivas procedentes de todos los sentidos (ahora no sólo los procedentes de la visión) no alcanzan las áreas emocionales el mundo pierde todo su sentido

emocional provocando dichos sentimientos de muerte y desolación. Pero aunque este síndrome es extremadamente raro, existen otros casos que sí son más frecuentes y que se van a examinar a continuación.

Los *síndromes de irrealización* y de *despersonalización*

En la práctica se presentan mucho más a menudo otro tipo de síndromes muy relacionados con los dos anteriores denominados *síndromes de irrealización* y de *despersonalización*; en el primero de ellos domina al paciente una sensación de que el mundo es *irreal* pareciendo como un sueño mientras que en el segundo de ellos se vive una vivencia del tipo *yo no soy yo*, con la sensación de ser un zombie. Como señala Vallejo Ruiloba [8] citando a su vez a Mellor, estos síndromes se caracterizan por: a) ser fenómenos subjetivos de la experiencia de uno mismo y del entorno, aparece una experiencia de cambio caracterizada por un sentimiento de extrañeza o de irrealidad, c) la experiencia es displacentera, d) va acompañada de otras alteraciones de las funciones mentales y e) se preserva el *insight*.

Ramachandran [6], siguiendo a Martin Roth, Mauricio Sierra y Germán Barríos, considera que estos síndromes podría tener un origen adaptativo consistente en *hacerse el muerto* del mismo modo que lleva a cabo la comadreja y otras especies ante la presencia cercana de un depredador. Con ello, la comadreja *desconecta* lo emocional del resto de las actividades corporales perdiendo todo el tono muscular y presentándose al exterior como un verdadero cadáver. Con esto conseguiría evitar al depredador dado que éstos no comen carroña y buscaría otra presa viva.

Siguiendo este razonamiento, el autor expone que este tipo de mecanismo se reproducirían en el ser humano en dos ocasiones distintas: a) en casos en los que hace falta tener *nervios de acero* para lo cual el aspecto emocional se reduciría al máximo ante la responsabilidad de ejecutar una acción extremadamente importante y b) en los casos en los que se produce una gravísima agresión o de peligro.

Si ahora se desciende al nivel neurológico intentando buscar una explicación de estos fenómenos se puede apreciar que ante casos de extrema gravedad para la integridad personal, el cingulado anterior genera una alerta extrema lo cual inhibiría la amígdala y demás centros límbicos emocionales suprimiendo temporalmente las emociones potencialmente incapacitantes como el miedo y la ansiedad .

Cognición emocional

Toda la evidencia planteada hasta este momento y que nos suministra la Neurología y la Psiquiatría, apunta a una visión holista e integradora de la cognición que pretende dar explicación real del ente humano de verdad, no del modificado y adaptado a los modelos teóricos de laboratorio. La realidad mostrada es que toda percepción se encuentra, a los pocos instantes, teñida de emoción por lo que, hablando en propiedad, debería hablarse del continuo *cognición-emoción*

adquiriendo pleno sentido las expresiones tan en boga como *inteligencia-emocional*.

Las implicaciones que de ello se deriva van mucho más allá del mero campo de las Neurociencias y la Filosofía de la mente. Un ser con una amplia y rica vida emocional tendrá unas cogniciones que le permitirán elaborar la realidad que le rodea de forma mucho más rica, elaborada y flexible de modo que ello le asegurará mayores probabilidades de éxito en cualquiera de los principales planos de su vida, en el ámbito profesional, personal y social. A *sensu contrario*, una deficiencia en los aspectos emocionales generará cogniciones deficientes, inadaptaciones que conducirán a fracasos de diversa índole y de distinta naturaleza , dicho con otras palabras, su procesamiento de la información será muy deficiente.

Referencias bibliográficas

[1] Bauer, R. M. (1984), 'Autonomic recognition of names and faces in prosopagnosia: a neuropsychological application of the guilty knowledge test', *Neuropsychologia* 22, pp. 457-69.
[2] Berrios G. E. y Luque R. (1995), 'Cotard´s Syndrome: analysis of 100 cases', *Acta Psychiatr Scand* 91(3), pp.185-8.
[3] Capgras, J. y Reboul-Lachaux, J. (1923), 'Illusion des sosies dans un délire systématisé chronique', *Bulletin de la Société Clinique de Médicine Mentale* 2, pp. 6-16.
[4] Ellis, H. D. y Young, A. W. (1990), 'Accounting for delusional misidentifications', *British Journal Psychiatry* 157, pp. 239-48.
[5] Hirstein, W. y Ramachandran, V. S. (1997), 'Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons', *Proceedings Royal Society London B Biological Science* 264, pp. 437-44.
[6] Ramachandran, V. (2003), *The Emerging Mind*, London, Profile books Ltd.
[7] Searle, J. (1994), *Mentes, cerebros y ciencia*, Madrid, Cátedra.
[8] Vallejo Ruiloba, J. (2000), *Introducción a la psicopatología y la psiquiatría*, Barcelona, Masson.

Dogmatism analysed

Mireia López Amo
Universitat de Girona, LOGOS
mireialopez.amo@gmail.com

The aim of this paper is to show that if Pryor does not pose externalist constrains on justification, his dogmatist view about perceptual justification is liberal to the implausible extent of allowing wishful thinking to count as justification. But if he includes such externalist conditions it is doubtful that his position can be qualified as internalist as it is his intention.

Pryor's Dogmatism

James Pryor presents his dogmatist view about perceptual justification as follows:

> The dogmatist about perceptual justification says that when it perceptually seems to you as if *p* is the case, you have a kind of justification for believing *p* that does not presuppose or rest on your justification for anything else, which could be cited in an argument (even an ampliative argument) for *p*. To have this justification for believing *p*, you need only have an experience that represents *p* as being the case. No further awareness or reflection or background beliefs are required. (Pryor 2000, p. 519)

Pryor's dogmatism gives an impulse to the Moorean antisceptical stance. It entails that subject's experiences alone justify a large class of perceptual beliefs, and it recovers Moore's basic idea that one can be justified in believing certain propositions without giving non circular evidence in their support.

Pryor's argument for dogmatism consists in showing that a subject can be immediately justified. Call a subject immediately justified in believing *p* just in case he is justified in believing *p*, and this justification does not presuppose or rest on any evidence or justification he has for believing other propositions. This allows perception alone to count as a source of justification and that is precisely dogmatism's point.

According to Pryor, it is intuitively natural to think that there is immediate justification. He claims that for propositions as <there are hands> one typically thinks that the mere fact of having the experience as of there being hands justify them, no further evidence or justified beliefs are needed. On Pryor's view, immediate justification for a belief does not require that this belief is infallible and indubitable. Moreover, he claims that immediately justified beliefs are neither self-evident nor self-justified nor autonomous beliefs. Also, he contends that the kind of justification immediately obtained is just *prima facie* justification; that is, justification that can be undermined by contrary evidence.

Pryor takes these features of immediate justification, as he understands it, to be sufficient to evade objections that have been mounted against traditional versions of foundationalism. Therefore, given the intuitive plausibility and the lack of extant criticisms to the thesis of immediate justification, this thesis has to be true.

However, even if a large class of propositions can be immediately justified, Pryor claims that it is not plausible to consider that every perceptual proposition we believe is immediately justified. We are immediately justified to believe perceptually basic propositions: propositions whose content we seem to perceive not by perceiving the content of other propositions, but by the mere deliverances of our experiences.

Pryor contends that a subject already possesses immediate justification to believe in a perceptually basic proposition merely by being presented with the content of that belief in the experience and by coming to hold that belief in the absence of countervailing evidence. In other words, it seems that subject's beliefs are immediately justified simply by taking the representational content of his experience at face value. However, being justified seems to require, both in the externalist and the internalist sense, more than taking content at face value.

Against the backdrop of his dogmatist account, Pryor now argues that he is in the position to resist pervasive scepticism. On Pryor's analysis, sceptical arguments proceed from the assumption that for a subject S to obtain perceptual justification to believe that *p*, it is required to have antecedent (non-question-begging) justification to believe that the sceptical hypotheses do not obtain. Yet, if the dogmatist account is correct, then one can obtain perceptual justification without being antecedently justified in believing that the sceptical hypotheses do not obtain.

The Non-Conceptual Nature of Perceptual Content

As I shall now argue, Pryor's dogmatist conception is committed to the thesis that there is something given in perception. This thesis consists in thinking that the mere contact with the external world provides some raw informational material to us. What is given in perception is ordinarily conceived as neither propositional nor conceptual in character because it is difficult to understand how what is given can obtain representational capacities and conceptual structure by itself.

On the assumption that there is nothing given in perception, perception is the result of interpreting sensory inputs in the light of other justified beliefs. Then the subject's perceptual beliefs are mediately justified. A subject is mediately justified in believing *p* just in case he is justified in believing *p*, and this justification rest in part in the justification he has for believing other supporting propositions. That is, if some beliefs are involved in the obtaining process of perception, they should be antecedently justified for the subject to get justification on their basis. Therefore, perceptual beliefs are mediately justified.

Consequently, if Pryor holds that one can be immediately justified on the basis of perception, he is accepting that something is given in perception. What is given is precisely what allows perceptual beliefs to be justified without requiring that this justification depends on the justification the subject has for other beliefs. On Pryor's view, perception is what allows for immediate justification, and this means that Pryor conceives perception as characteristically given.

Attending to the properties of the given, Pryor seems to be committed to both *i*) a non-propositional view about perceptual content and *ii*) the non-conceptual character of perceptual content. Pryor explicitly rejects *i*) and says that the content of experience represents the world as being in a certain way. However, given that he needs to preserve some trace of the direct character of perception, he cannot avoid accepting its non-conceptual constitution.

Basic Beliefs

Pryor calls himself *modest foundationalist*. It is "modest" in the sense that he does not attribute to immediately justified beliefs all those properties –infallibility, indubtability, self-evidentiality, etc- that classical foundationalism attributes to them. However, as a foundationalist, he cannot evade recognizing the existence of basic beliefs that are not ordinarily justified by other beliefs.

Classical foundationalism characterizes basic beliefs as beliefs whose justification does not rest on the justification for any other beliefs. This suffices for its purposes of finding a candidate to stop the regress of justifications. It allows the existence of another source of justification, perceptual experience that does not seem to require further justification. Basic beliefs thus characterized seem to be sufficient also for Pryor's purposes against pervasive scepticism. Yet Pryor's conception of basic beliefs seems to go further than classical foundationalist conception.

On Pryor's view, basic beliefs are beliefs that are immediately justified. A belief is immediately justified, not merely when its justification does not rest on the justification for any other beliefs, but also when its justification does not presuppose or rest on any evidence the subject has for believing other propositions. On his view, the justification for basic beliefs does not rest on anything else that could be cited as a premise in an argument for them.

Suppose I believe that these walls are red. What can justify me in believing this proposition is the evidence I have of my experience; that is, my awareness of me having that experience. However, my belief that these walls are red can still be justified without receiving its justification from other beliefs. Consequently, my belief could be basic in the sense specified by classical foundationalism but not in Pryor's sense.

Justification and wishful thinking

Now, I will argue that Pryor's commitment to the content of perception being non-conceptual and his characterization of basic beliefs conflicts with the thesis that justification requires an internalist condition about the subject being aware of his justifiers.

The relation of justification holds between contents. This relation is considered rational only if consists in some semantic relation. If one of its relata is non-conceptual, no semantic relation can be obtained. Therefore, if one insists in talking about justification in this case, he cannot be talking of a rational relation. Pryor thinks that the content of perception justifies the subject's belief but maintains that this content is non-conceptual. Consequently, his conception of justification cannot be characterized as rational.

What other relation can hold between the non-conceptual content of experience and perceptual beliefs? The relation can be merely causal: having certain experience as of *p* causes the subject's belief that *p*. Thus, Pryor's can be taken as conceiving justification as the mere result of the occurrent experience causing subject's beliefs. Therefore, no condition about the subject being aware of his experience seems to be required for justification.

Let us now see how Pryor's commitment to mentalism can be made explicit out of his characterization of basic beliefs. An internalist about justification claims that what justify our beliefs are facts internal to the subject. However, there are two ways to describe what is internal: *i*) metaphysically and *ii*) epistemically. Mentalism arises from *i*) and the view that derives from *ii*) is known as accessibilism. Thus we get two versions of internalism:

> **Mentalism**: what justifies any subject's beliefs are his mental states.
>
> **Accessibilism**: what justifies any subject's beliefs must be reflectively accessible to the subject.

Pryor characterizes basic beliefs as perceptual beliefs that are immediately justified. We have seen that justification for basic beliefs does not presuppose or rest either on any other justified beliefs or on subject's evidence for other propositions, since they receive justification from perception alone. Consequently, it can be said that a subject's belief that *p* is basic if the mere having an experience as of *p* suffices for being justified in believing that *p*. But having an experience as of *p* is no more than being in a mental state representing the world as of *p*. Therefore, a subject's belief that *p* is basic if the subject's being in a mental state representing the world as of *p* suffices for being justified in believing that *p*.

Note that if being in a certain mental state suffices for being justified, the subject's mental states determines whether the subject has justification and this claim appears to be just a mere a reformulation of mentalism.

Not all mental states are accessible to the awareness of the subject. Consequently, if mentalism just claims that what justifies any subject's belief are

his mental states, it does not seem to be required that the subject also needs to be aware of the content of his states. Mentalism, indeed, states that is the mere occurrence of subject's mental state that gives the subject perceptual justification.

From Pryor's non-conceptual view about perceptual content I have derived that perceptual justification is the mere result of occurrent experience causing subject's beliefs. I have also shown that his characterization of basic beliefs leads to the thesis that the mere occurrence of the subject's mental state gives perceptual justification to the subject. I will argue now that if these theses correctly describe sufficient conditions for perceptual justification, then the doxastic output of any cognitive mechanism of belief formation could count as justified, even if only *prima facie*.

Suppose that a subject is in a mental state that represents the world as being *a certain way W* because he strongly desires the world being *W*. Then, he comes to believe that the world is *W* on the basis of that experience. According to Pryor, the subject is justified in believing that the world is *W* because his belief has been caused by his experience as of the world being *W* and it is the mere occurrence of that experience that provides justification.

Although under Pryor's assumptions there is justification in this case, intuitively we do not think that such subject is justified. How can a subject be justified in believing that the world is *W* by the mere fact of desiring the world being that way (even if we consider that he is only *prima facie* justified)? How can a subject be justified by mere wishful thinking?

We think of justification as constrained by some requirements, if not by internalist conditions requiring subject's awareness of his evidence, at least constrained by some external conditions concerning the correctness of the belief's forming process. These conditions appear as useful to rule out the counterexamples that an excessively liberal view is subject to.

As we have seen, Pryor is reluctant to take as necessary for justification the subject's awareness of his evidence. Moreover, he overtly accepts that the subject does not need to positively assume that no defeaters are in place to be justified. Therefore, if he aims to preclude taking wishful thinking as a source of justification, he has to be disposed to accept some external condition constraining justification.

My objective has been to show that Pryor's view is too liberal to the extent of allowing wishful thinking to count as justified. However, if he wants to avoid this undesirable consequence he has to pose some external constrain on justification, since he implicitly rejects including any internalist condition about subject's awareness. But this makes his view about perceptual justification an externalist one while he calls himself internalist.

References

Audi, R. (1988), 'Foundationalism, Coherentism and Epistemological Dogmatism', *Philosophical Perspectives*, vol. 2, Epistemology, pp. 407-42.

Bergmann, M. (2006), *Justification without Awareness*, Oxford, Oxford University Press.

Brewer, B. (1999), *Perception and Reason,* Oxford, Oxford University Press.

Coliva, A. (2009), 'Moore's Proof, liberals and conservatives. Is there a third way?', in A. Coliva (ed.), *Mind, Meaning and Knowledge. Themes from the Philosophy of Crispin Wright*, Oxford University Press, *forthcoming*.

Heck, R. (2000), 'Nonconceptual Content and the "Space of Reasons"', *Philosophical Review* 109.4, pp. 483–523.

McDowell, J. (1994), *Mind and World*, Boston, Harvard University Press.

Pryor, J. (2000), 'The skeptic and the dogmatist", *Noûs* 34, pp. 517-49.

— (2001), 'Highlights of Recent Epistemology', *British Journal for the Philosophy of Science* 52, pp. 95–124.

— (2004), 'What's wrong with Moore's argument?', *Philosophical Issues* 14, pp. 349-78.

— (2005), 'There is immediate justification', in M. Steup and E. Sosa (eds.), *Contemporary Debates in Epistemology*, Oxford, Blackwell, pp. 181-202.

Steup, M. (2004), 'Internalist Reliabilism', *Philosophical Issues* 14, Epistemology, pp. 403-25.

Las representaciones y la distinción sintaxis – semántica

Camilo Andrés Ordóñez Pinilla
Universidad Nacional de Colombia
camilo.ordonez@gmail.com

La Ciencia Cognitiva representacionalista postula las siguientes características como rasgos de los que debe dar cuenta una explicación acerca de qué son las representaciones mentales: **(1)** los procesos mentales que tratan con representaciones son computacionales (o, en una versión más moderada, pueden ser descritos computacionalmente); y **(2)** las representaciones tienen contenidos acerca del ambiente. En otras palabras, la Ciencia Cognitiva representacionalista considera que una teoría de la representación debe explicar la *manipulabilidad computacional* e *intencionalidad* de las representaciones.

En C3 [Cussins (1994)], Adrian Cussins propone entender las representaciones como entidades compuestas por un *vehículo* y un *contenido*. Esta caracterización de las representaciones nos da una idea de cómo explicar su manipulabilidad computacional e intencionalidad, en tanto las propiedades del *vehículo* de la representación son de naturaleza *computacional* y que las propiedades del *contenido* de la representación son de naturaleza *semántica*. De esta manera, una explicación de qué son las representaciones necesitaría de una buena caracterización las propiedades computacionales y de las propiedades semánticas.

Tradicionalmente la noción de 'propiedad computacional' se ha entendido a través de la noción de 'algoritmo': una propiedad es computacional si y sólo si puede ser capturada en la especificación de un algoritmo. Los algoritmos se entienden como conjuntos de instrucciones que son *efectivos* (i.e. que garantizan el éxito en una tarea) y *semánticamente insensibles* (i.e. que pueden seguirse sin tener en cuenta, o ser afectados por, los valores semánticos de los ítems sobre los que opera, como la verdad o el significado). A su vez, la teoría de la computación ha considerado que tales características de los algoritmos dependen de entenderlos como atendiendo a propiedades puramente formales o *sintácticas*. Así, tenemos que la manera tradicional de entender las propiedades computacionales es entenderlas como propiedades de naturaleza sintáctica.

Esto nos lleva a que para construir una teoría de la representación es necesario poder caracterizar, dada una representación, cuáles propiedades pueden considerarse sintácticas y cuáles semánticas. La concepción tradicional de las propiedades semánticas es entenderlas como las propiedades de las representaciones que se determinan en virtud de las relaciones de ellas con aquello que representan. A partir de esto es posible construir una noción de las propiedades sintácticas, desde un punto de vista negativo, diciendo que son las propiedades de las representaciones que no se determinan en virtud de las

relaciones de ellas con lo que representan. La versión positiva de esta caracterización consiste en entender las propiedades sintácticas como aquellas que se determinan en virtud de características formales (i.e. que atienden sólo a la forma) de las estructuras que componen la representación.

Pero, tales caracterizaciones tradicionales de las propiedades sintácticas y semánticas son problemáticas. Un primer asunto problemático acerca de las propiedades sintácticas es que es difícil encontrar una explicación precisa de qué se quiere dar a entender al hablar de las propiedades que depende sólo de la *forma* de las estructuras que componen la representación. Por lo general las explicaciones de qué es lo sintáctico se detienen en una repuesta en la que se dice que lo sintáctico es lo que depende de la forma y a su vez, que cuando hablamos de 'forma' no se tiene en cuenta el significado. Pero, aún sigue siendo válido y casi necesario preguntarse por a qué tipo de cosas nos referimos cuando hablamos de la 'forma' de una estructura representacional. La intuición más natural que parece generarse a partir de esto es que las únicas características con las que contamos para dar cuenta de qué es la 'forma' de una estructura representacional son sus características geométricas. Si bien al decir esto, todo queda por ser explicado.

Pero, de aquí se sigue un problema fundamental para la concepción clásica, en tanto es posible encontrar un caso en el que la distinción sintaxis – semántica se desdibujaría: según el análisis tradicional, entre lo sintáctico y lo semántico se da una relación de exclusión (i.e. lo semántico no es sintáctico y lo sintáctico no es semántico). Además, según se ha afirmado anteriormente, dada una estructura representacional, las propiedades sintácticas de tal estructura son aquellas que están determinadas en virtud de sus propiedades geométricas, mientras que sus propiedades semánticas serían las que se determinan en virtud de propiedades de *lo representado* (vg. el mundo del lenguaje natural es el mundo de la experiencia, y por tanto, son las relaciones con el mundo de la experiencia las que determinan las propiedades semánticas de los ítems lingüísticos del lenguaje natural). Ahora bien, cuando un geómetra realiza su actividad, el dominio de lo representado está dado por el *mundo* de la Geometría. De esta manera, a la pregunta por cuáles son las propiedades semánticas de una representación de ese *mundo* tendría que responderse: las propiedades de tal representación que se determinen en virtud de propiedades del *mundo* de la Geometría; que evidentemente serían propiedades que podríamos llamar propiedades geométricas. Y, de la misma manera, a la pregunta por cuáles son las propiedades sintácticas de una representación de ese mundo, tendríamos que decir que serían propiedades geométricas. Pero, esto genera un problema para la concepción tradicional de la distinción entre sintaxis y semántica, puesto que nos deja sin una manera de entender en qué sentido las propiedades sintácticas son diferentes de las propiedades semánticas. Si bien esto no puede contar como un argumento contundente en contra de la concepción clásica, es suficiente para mostrar la necesidad de pensar en una manera diferente de capturar la distinción entre sintaxis y semántica, en tanto la concepción clásica no puede considerarse como brindando una explicación general de tal distinción.

De esta manera, para poder tener una teoría de la representación se deberían aceptar dos retos fundamentales: primero, precisar la intuición de que cuando se

habla de 'forma', tal noción se entiende como una noción geométrica. Segundo, construir la distinción entre sintaxis y semántica de una manera que no implique una distinción entre propiedades que dependen de lo representado y propiedades que no dependen de lo representado, para poder evitar el incómodo caso del geómetra. El establecimiento de estos retos supone que mi idea es que lo sintáctico puede caracterizarse de una manera geométrica, en tanto no veo más candidatos plausibles, y que por ende, para evitar el caso del geómetra, no necesitamos rechazar esta idea, sino la concepción de que la semántica y sólo la semántica es el nivel que depende intrínsecamente, y se relaciona fundamentalmente con, *lo representado*.

Acerca del primero reto, quisiera pensar que las estructuras que componen la representación deben entenderse como análogos a construcciones que se realizan a partir de unos ítems primitivos y unas reglas de construcción, donde ambas cosas están definidas de manera geométrica. La 'Geometría' parte de unas nociones idealizadas de punto y/o línea, para, entendiendo esto en un sentido no sofisticado, construir 'cosas' con puntos y líneas. Un punto, una línea, y cualquier otra cosa que pueda ser construida con puntos y líneas puede considerarse un *ítem geométrico*. Es en este sentido que los ítems primitivos de las construcciones que estoy caracterizando estarían definidos de manera geométrica. Por su parte, las *reglas* pueden entenderse como las reglas de los sistemas formales que nos dicen cuando es correcto poner dos ítems o dos conjuntos de ítems juntos y cuando es incorrecto hacerlo. Estas reglas contarían como reglas geométricas, en tanto están definidas para ítems geométricos y para ser usadas sólo requieren habilidades de reconocimiento de tales ítems y una especie de nociones cuasi-espaciales de 'estar al lado de'.

Tal caracterización enfrenta un problema fundamental: si entendemos la noción de 'forma' en términos geométricos, la identidad de las estructuras que componen la representación va a depender de relaciones geométricas. Pero, no es deseable que propiedades como 'ser más grande qué' determinen la identidad de los elementos de una sintaxis. En este caso queremos que dos cosas puedan constar como una instancia de un ítem sintáctico, amén de que una sea más grande que otra (vg. si una letra proposicional se escribe más grande que otra en dos líneas de una demostración, no por eso deberían contar como diferentes). Para subsanar este problema quiero proponer que la identidad de tales construcciones geométricas depende, no de sus mapas geométricos, sino de sus mapas topográficos. En líneas generales un mapa topográfico puede caracterizarse de la siguiente manera: un mapa es una representación gráfica de los objetos que ocupan un cierto espacio, en el que están capturadas relaciones espaciales entre ellos. Las relaciones espaciales que se encuentran en los mapas se dividen en dos: de distancia (i.e. que tan lejos está un objeto de otro) y de ubicación (i.e. en que dirección y posición relativas se encuentra un objeto con relación a otro). La noción de mapa topográfico que quiero aquí es la que propone Paul Churchland para explicar la manera en que está representado el espacio en el cerebro [Churchland (1986), pp. 81-83]. En los *mapas topográficos*, se representan los objetos sin respetar las relaciones de distancia, sino únicamente las relaciones de ubicación, que Churchland llama

relaciones vecinales. De esta manera, dentro de un mapa topográfico, dos objetos pueden contar como el mismo, aún si uno es más grande que el otro, dado que sólo es necesario que en ambos se den el mismo de relaciones de ubicación relativas, entre los ítems que lo componen.

En cuanto al segundo reto quisiera proponer una nueva manera de entender la noción de 'semántica'. Postulamos las propiedades semánticas para tener una explicación de la intencionalidad de las representaciones. Esto implica que todo intento por precisar qué son las propiedades semánticas debe hacerlo de una manera tal que permita una relación entre la representación y lo que representan. Pero, el caso del geómetra nos muestra que tal relación no debe ser lo característico de lo semántico. Así, es posible proponer que si bien las propiedades semánticas no son propiedades *determinadas* en virtud de relaciones con lo que representan, deben definirse en términos de que a través de ellas el ambiente está disponible a un ser cognitivo (i.e. intencionalidad). Pero, esto debe precisarse, en tanto en la definición de 'propiedad sintáctica' se apeló a la noción de 'algoritmo', entendiendo este último en términos de instrucciones semánticamente insensibles. Entonces, si es correcto postular *actividades* semánticamente insensibles, es decir, que surgen de seguir instrucciones semánticamente insensibles, es necesario tomarse seriamente la idea de que las propiedades sintácticas pueden contar como guías de la actividad, y por tanto, como *maneras en las que el ambiente está disponible para un ser cognitivo*, en tanto todas las actividades deben ser, en algún grado, sensibles a características del ambiente en el que se ejecuta la actividad.

Así, el análisis propuesto implica que para distinguir entre sintaxis y semántica es necesario caracterizar las propiedades semánticas de una manera tal que ellas presentan el ambiente de una determinada manera, pero que al decir esto no se excluya la posibilidad de que las propiedades sintácticas se puedan concebir como maneras en las que el ambiente está disponible a un ser cognitivo. La idea es, entonces, que la distinción entre sintaxis y semántica es una distinción entre maneras en que el ambiente está disponible para un ser cognitivo.

Para poder lograr tal caracterización, me gustaría proponer que las maneras en las que el ambiente está disponible para un ser cognitivo pueden ser entendidas como construcciones realizadas a partir de ítems que se '*refieren'* a los constituyentes de los estados del ambiente. Las propiedades sintácticas son propiedades que *dependen* de la construcción y las propiedades semánticas son propiedades que dependen de una *interpretación* de tal construcción. Formalmente, se dice que una propiedad Φ *depende* de su construcción *P*, si y sólo si el hecho de que *R* instancie Φ es una *consecuencia necesaria* —no *contingente*— de un paso o un conjunto de pasos de *P*. Por su parte, las propiedades semánticas no están dadas *en,* ni tampoco dadas por, el proceso de la construcción, sino que, están determinadas por tomar las propiedades geométricas de una construcción y darles un *valor*, una *asignación*, en otra estructura.

Referencias bibliográficas

Churchland, P. (1986), 'Some Reductive Strategies in Cognitive Neurobiology', *Mind*, Vol. XCV, No. 379, pp. 279-309.

Cussins, A. (1994), 'La construcción conexionista de conceptos', en Boden, M. (ed.), *Filosofía de la Inteligencia Artificial,* México D.F., Fondo de Cultura Económica, pp. 409-87.

Conocimiento, discriminabilidad y acceso al contenido representacional*

Manuel Pérez Otero
Universidad de Barcelona, LOGOS
perez.otero@ub.edu

1. Intuiciones y teorías epistemológicas diversas indican que algún requisito acerca de las capacidades discriminatorias es necesario para que haya conocimiento. Se trataría de alguna versión convenientemente restringida de lo que en una primera aproximación podría formularse así:

> *Postulado de Discriminabilidad* (**PD**): Si S sabe que *p*, entonces S tiene capacidad para discriminar entre el caso de que *p* y otras alternativas relevantes.

Hagamos dos clarificaciones preliminares. Primero, en relación con las alternativas relevantes: es plausible considerar que frecuentemente la alternativa relevante al caso de que *p* es, por defecto, el caso de que no *p*. En segundo lugar, conviene concebir la capacidad a la que se alude en un sentido relativamente débil –o bien, alternativamente, proponer otra versión del postulado, con el mismo propósito– para evitar que el postulado parezca demasiado rápidamente descartable, a la vista de posibles escenarios como los que se utilizan para construir argumentos escépticos cartesianos sobre el conocimiento perceptivo.

Con las matizaciones que se han hecho, (**PD**) parece expresar un principio verdadero y considerablemente básico sobre el conocimiento. Asumiendo eso, cabe –de forma simplificada– distinguir dos opciones alternativas:

a) tomar efectivamente (**PD**) como principio epistemológico básico;
b) tratar de explicar o elucidar (**PD**), ofreciendo para ello alguna otra idea o principio epistemológico que (en general, o quizá al menos para estudiar ciertas cuestiones) se propone como más básico.

Me inclino por la opción (b). Propongo que (**PD**), como sucede con otros principios, deriva de una idea similar, pero algo diferente, también integrante de nuestro concepto del conocer. Es el postulado de que el conocimiento constituye una garantía falible contra el error:

> *Conocimiento como Garantía Contra el Error* (**GCE**): Si S sabe que *p*, entonces dicho conocimiento le confiere cierta garantía (posiblemente falible) contra el riesgo de error.

---

* Este trabajo forma parte del Proyecto de Investigación "Discriminabilidad: representación, creencia y escepticismo", subvencionado por el Ministerio de Ciencia e Innovación (FFI2008-06164-C02-01). Algunas ideas aquí presentadas surgieron a partir de debates en el grupo de lectura Logos sobre *Transparencia del contenido y autoconocimiento*, que coordiné durante el curso 2007-08. Agradezco a sus integrantes su participación, muy especialmente a Ekain Garmendia, con quien he discutido extensamente sobre estos temas.

Al formular (**GCE**), podría también ponerse "garantía fiable pero falible contra el error". Efectivamente, pretendo que la garantía mencionada conlleva cierta fiabilidad. Si prescindo de hacerlo explícito es porque entiendo que 'garantía' ya transmite claramente ese rasgo. Por otro lado, invoco un sentido de 'garantía' compatible con la falibilidad. Puesto que en su sentido usual, 'garantía' no conlleva falibilidad –e incluso puede pensarse que es incompatible con ella– conviene hacer esta observación, e incorporar explícitamente en (**GCE**) el calificativo 'falible'.

Una clarificación sobre el tipo de error mencionado en (**GCE**) permite ver la conexión entre (**GCE**) y (**PD**); es el error de tomar el caso de que *p* por alguna de las alternativas relevantes (o viceversa). Por ejemplo, saber que *p* nos confiere cierta garantía de que al creer así que *p* no estamos sin embargo en el caso de que no *p*. En ese sentido, conocer que *p* implica poder discriminar si es el caso que *p* o es el caso que no *p*. Dicho de otro modo: la discriminabilidad en cuestión, postulada por (**PD**), es la que viene implicada por (**GCE**).

2. La estrategia propuesta, consistente en ofrecer en la explicación-elucidación de un principio epistemológico suficientemente básico, (**PD**), otro principio, (**GCE**), que puede considerarse más fundamental, tiene aplicaciones concretas a debates contemporáneos. En el resto de esta comunicación abordo uno de esos debates: el presunto conflicto entre el externismo intencional y la tesis del Acceso Privilegiado (según la cual cada sujeto conoce por introspección cuáles son los contenidos de sus pensamientos). Mi posición será compatibilista: no hay conflicto real entre ambas tesis; las versiones apropiadas de cada una de ellas son consistentes entre sí. Usaré (**GCE**) para defender el compatibilismo, clarificando algunos puntos y tratando de desterrar algunos errores.

La tensión entre externismo y Acceso Privilegiado tiene dos vertientes diferentes. Sólo una nos concierne aquí. Conviene distinguirlas. El primer problema para el compatibilismo lo presentó McKinsey (1991). Ese inconveniente consiste en lo siguiente. El externismo es una teoría a la que se llega mediante la investigación teórica filosófica. Supuestamente tendría como consecuencia, por ejemplo, que para pensar sobre el agua es necesario haber estado conectado, directa o indirectamente con agua. Por otro lado, según Acceso Privilegiado, accedemos por introspección al contenido de nuestros pensamientos. Reuniendo esos elementos, tendríamos que un sujeto que sabe que está pensando que el agua es saludable puede concluir –usando meramente la introspección y el tipo de investigación teórica característica de la filosofía– que existe (o ha existido) el agua. Eso parece insostenible. Pero no es éste el reto incompatibilista que pretendemos explorar ahora.[1]

2.i. El otro problema lo presentó Burge (1988), proponiendo para él una solución compatibilista; aunque seguramente ha sido Boghossian (1989), desde posiciones

[1] Me he ocupado de ese problema en Pérez Otero (2004a y 2004b).

incompatibilistas, quien más lo ha popularizado. Resumamos telegráficamente el experimento mental que mejor ilustra el externismo intencional y contribuye a caracterizarlo (cf. Putnam 1973 y Burge 1979). Supongamos que Oscar y un duplicado suyo, Bi-Oscar, habitan, respectivamente, nuestra Tierra y una Tierra Gemela, Bi-Tierra, en una época anterior al desarrollo de la teoría química. Utilizan de manera completamente análoga el término 'agua', asociando a él exactamente las mismas creencias consideradas desde un punto de vista subjetivo. En la Bi-Tierra, sin embargo, la referencia de 'agua' no es el agua, sino un líquido de apariencia indistinguible pero estructura atómica diferente, XYZ, al que podemos llamar 'bi-agua'. El externismo lingüístico inicialmente propugnado por Putnam establece que Oscar y Bi-Oscar usan 'agua' con significados diferentes. El externismo intencional defendido por Burge añade que también difieren típicamente sus creencias cuando, por ejemplo, creen lo que ambos expresarían diciendo 'El agua es saludable'.

Veamos ahora cuál es el problema. Supongamos, en efecto, que Oscar y Bi-Oscar profieren sinceramente 'El agua es saludable'. Usando mayúsculas para referirnos a los correspondientes conceptos, Oscar cree el contenido proposicional EL AGUA ES SALUDABLE, mientras que Bi-Oscar cree el contenido proposicional LA BI-AGUA ES SALUDABLE, distinto del anterior conforme al externismo intencional. Por otro lado, Acceso Privilegiado implica que Oscar sabe, por mera introspección, qué pensamiento constituye el contenido proposicional de su creencia. Es decir, implica que el enunciado (*), pronunciado por Oscar, expresaría una verdad que Oscar conoce por introspección:

(*) Creo que el agua es saludable

Según Boghossian es problemático armonizar ambos resultados: Oscar no sabe que cree que el agua es saludable porque no podría distinguir por introspección si es EL AGUA ES SALUDABLE o bien LA BI-AGUA ES SALUDABLE el objeto de su creencia (cf. Boghossian 1989).

2.ii. Si –como he sugerido– la plausibilidad de (**PD**), del que depende la existencia del problema, deriva de (**GCE**), entonces el problema se disuelve. No hay impedimento a atribuir a Oscar conocimiento por introspección de (*) porque no hay riesgo de que aunque le parezca creer EL AGUA ES SALUDABLE esté creyendo en realidad LA BI-AGUA ES SALUDABLE (ni viceversa). El externismo impone condiciones para poder tener el pensamiento LA BI-AGUA ES SALUDABLE que Oscar no satisface, pues no habita el entorno apropiado.

Los filósofos que tratan estos temas han complicado las cosas, complementando la trama del experimento mental original con unos cuantos viajes. Sin saberlo, Oscar viaja a la Bi-Tierra y permanece allí tiempo suficiente como para, aparentemente, adquirir el nuevo concepto: BI-AGUA. Luego regresa a la Tierra, por un tiempo también prolongado. Y así sucesivamente, en varias ocasiones posteriores. Esta nueva versión, de *cambios lentos* [*slow switchs*], según suele llamársele, aporta una modificación significativa. En el ejemplo inicial el argumento incompatibilista podría ser fácilmente refutado alegando que la

circunstancia de estar creyendo el otro contenido proposicional, LA BI-AGUA ES SALUDABLE, no es una alternativa relevante (lo cual puede entenderse en términos de *cercanía* de los mundos posibles en que se tendría esa otra creencia). Con el nuevo escenario esa réplica parece incorrecta.[2]

Pero la solución compatibilista que hemos mencionado –basada en (**GCE**)– se aplica también en el nuevo escenario. Para ser precisos, hay dos tesis alternativas principales sobre lo que sucede cuando tienen lugar tales cambios lentos. Algunos asumen la tesis del *reemplazo* de conceptos: tras su primera estancia prolongada en la Bi-Tierra, por ejemplo, Oscar deja de poseer el concepto anterior, AGUA, y pasa a poseer sólo un nuevo concepto, diferente según las versiones: el concepto BI-AGUA (cf. Ludlow 1995), o quizás un concepto amalgama cuya extensión incluye el agua y la bi-agua, así como la extensión de JADE incluye la jadeíta y la nefrita (cf. Falvey 2003). Otros asumen como más plausible la *cohabitación*: Oscar no deja de poseer el concepto AGUA después de haber adquirido el concepto BI-AGUA (cf. Burge 1988, 1998).

Mi propuesta de que las capacidades discriminatorias exigidas por (**PD**) se entiendan conforme a (**GCE**) se aplica paradigmáticamente en los casos de reemplazo, eliminando por completo el problema: cuando está en la Tierra, no hay riesgo de error para Oscar al creer que está pensando EL AGUA ES SALUDABLE, porque si estuviera en la Bi-Tierra no se equivocaría (pues entonces creería que está pensando LA BI-AGUA ES SALUDABLE).

En las situaciones de cohabitación la propuesta también sería aplicable, pero de forma no tan obvia, teniendo en cuenta –además– que la cohabitación de conceptos suscita complicaciones adicionales (por limitaciones de espacio, no podemos abordar estos puntos).[3]

2.iii. El propio Boghossian contempla la posibilidad de una réplica a su argumento similar a la esbozada, dependiente de que –conforme a la tesis del reemplazo– en cada momento Oscar no puede tener el concepto apropiado que le permita excluir la hipótesis alternativa relevante. Pese a todo, aduce, Oscar debe poder excluir la alternativa relevante; si un sujeto no puede descartar que la moneda que tiene en la

---

[2] Los dos casos discurren paralelamente a dos ejemplos clásicos que ilustran bien la idea de *alternativa relevante* y su importancia para el conocimiento. En condiciones normales, ver un granero nos permite saber que estamos ante un granero, porque la posibilidad de que sea un decorado que parece un granero es remota y, por ello, no constituye una alternativa relevante. Pero si vemos un granero en un entorno donde una proporción importante de aparentes graneros son decorados, entonces la alternativa de estar ante uno de esos decorados es relevante y determina que no haya conocimiento. (Cf. Goldman 1976).

[3] Mi respuesta compatibilista es similar a la de Falvey y Owens (1994), que asumen la tesis del reemplazo. Sin embargo, estos autores adoptan una posición excesivamente defensiva, al proponer que el autoconocimiento del contenido no involucra capacidades discriminatorias, para lo cual caracterizan y rechazan una tesis que denominan "conocimiento introspectivo del contenido comparativo" (pp. 109-110); su caracterización es inapropiada, pues la tesis tiene entonces contraejemplos mucho más triviales que los que ellos presentan.

mano sea falsa (en un entorno en que proliferan monedas falsas) entonces no sabe que tiene una moneda auténtica, incluso si carece del concepto de moneda falsa (cf, Boghossian 1989, p. 14 y nota 12).

Pero Boghossian interpretaría incorrectamente la razón fundamental de que mediante un principio como (**GCE**) se bloquee su argumento. Por no tener el concepto alternativo requerido, *necesario para representarse canónicamente el contenido proposicional correspondiente*, no hay riesgo de error cuando Oscar considera su pensamiento. En el caso de la moneda falsa, lo importante no es tener o no el concepto de moneda falsa, sino que el sujeto tendrá un medio de representarse (perceptivamente) la moneda falsa, con el consiguiente riesgo de error, incompatible con (**GCE**).

Referencias bibliográficas

Boghossian, P. A. (1989), 'Content and Self-Knowledge', *Philosophical Topics* 17, pp. 5-26.

Burge, T. (1979), 'Individualism and the Mental', *Midwest Studies in Philosophy* 4, pp. 73-121.

— (1988), 'Individualism and Self-Knowledge', *Journal of Philosophy* 85, pp. 649-663.

— (1998), 'Memory and Self-Knowledge', en Ludlow, P. y Martin, N. (eds.), *Externalism and Self-Knowledge*, Stanford, CSLI Publications, pp. 351-70.

Falvey, K. (2003), 'Memory and Knowledge of Content', en Nuccetelli, S. (ed.), *Semantic Externalism and Self-Knowledge*, Cambridge, MIT Press, pp. 219-40.

Falvey, K. y Owens, J. (1994), 'Externalism, Self-Knowledge, and Skepticism', *Philosophical Review* 103, pp. 107-37.

Goldman, A. I. (1976), 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* 73, pp. 771-91.

Ludlow, P. (1995), 'Externalism, Self-Knowledge, and the Prevalence of Slow Switching', *Analysis* 55, pp. 157-59.

McKinsey, M. (1991), 'Anti-Individualism and Privileged Access', *Analysis* 51, pp. 9-16.

Pérez Otero, M. (2004a), 'Las consecuencias existenciales del externismo', *Análisis Filosófico* 24, pp. 29-58.

— (2004b), 'El externismo intencional ante la transparencia de las actitudes proposicionales', en A. Vicente, P. de la Fuente, C. Corredor, J. Barba y A. Marcos (eds.), *Actas del IV Congreso de la SLMCFE*, Valladolid, pp. 262-65.

Putnam, H. (1973), 'Meaning and Reference', *Journal of Philosophy* 70, pp. 699-711.

Lógica, pragmática y pragmatismo: El análisis de las actitudes cognitivas en C. S. Peirce*

Mª Uxía Rivas Monroy
Universidad de Santiago de Compostela
uxia.rivas@usc.es

Introducción

El interés de las reflexiones de Peirce sobre actitudes cognitivas, tales como el juicio, la aserción o la creencia radica, fundamentalmente, en la gran riqueza y variedad de elementos que tuvo en cuenta, mostrando la complejidad de estas acciones humanas que involucran signos, razonamiento y aspectos valorativos. Así, primeramente, el análisis peirceano presenta la proposición como el componente lógico-semiótico presente en el juicio, la aserción o la creencia. En segundo lugar, el aspecto pragmático se pone de relieve al entender Peirce estas actitudes cognitivas como acciones, que, en el caso de la creencia define explícitamente como hábito de conducta. Y en tercer lugar, destaca también la dimensión pragmatista, que vincula hechos y valores, y que involucra aspectos normativos, presentes principalmente en la aserción, al incluir la responsabilidad por parte del sujeto. El pensamiento de Peirce sobre estos temas se puede relacionar con dos líneas de investigación contemporáneas: a) la dimensión pragmática, de raíz austiniana, que permite reinterpretar sus ideas en la clave terminológica proporcionada por Austin en su doctrina de los actos de habla; y b) la dimensión pragmatista, que, reinterpretada por Putnam, invita a eliminar los dualismos en los análisis filosófico-lingüísticos, en concreto, el de hecho/valor, y que tiene en Peirce un predecesor a la vista de sus análisis del juicio y la aserción. De estos últimos Peirce destacará no sólo su naturaleza representativa (lógico-semiótica), sino también su vinculación a efectos y consecuencias reales, en consonancia clara con su pragmaticismo.

La dimensión lógica y pragmática de las actitudes cognitivas

El juicio, la aserción y la creencia son actitudes cognitivas que se apoyan en un componente esencialmente lógico-semiótico, que es la proposición, la cual funciona como la materia sobre la que los hablantes se posicionan de manera diversa, bien juzgándola, aseverándola o creyéndola. Cada una de estas actitudes cognitivas puede expresarse a través de diferentes actos de habla, con sus peculiaridades pragmáticas que establecen, en gran medida, la diferencia entre ellas.

* Este trabajo se realizó en el marco del proyecto de investigación HUM2006-04955/FISO del Ministerio de Ciencia y Teconología.

La proposición

En líneas generales, el análisis de la proposición de Peirce se mantiene fiel a la concepción tradicional que la analiza en términos de sujeto y predicado, incidiendo en su carácter informativo. Peirce destaca el carácter diádico de la proposición y se centra en su análisis semiótico, prestando atención a los tipos de signo que la forman, fundamentalmente, iconos e índices, aunque ambos tengan en la proposición el carácter de símbolos. Este análisis semiótico es una de las mayores aportaciones de Peirce a la comprensión de la proposición, así como su vinculación con los actos de habla de la aserción y el juicio.

Normalmente, Peirce analiza la proposición a partir de su concepción de la semiosis[1] y de sus clasificaciones de los signos. Así, define a la proposición como un símbolo, cuyo interpretante la representa como un icono de un índice del individuo nombrado (CP 2.329). En la proposición el sujeto es representado por un índice y el predicado por un icono. Un índice es un signo que es determinado por el objeto dinámico en virtud de una relación de proximidad o contigüidad con el mismo. Los nombres propios, los pronombres personales o los demostrativos son ejemplos muy claros de índices que muestran esa conexión directa con el objeto, de modo semejante a como una veleta muestra una relación directa con la dirección del viento, o una huella con el pie que la marcó. Una proposición necesita tener esa vinculación con un objeto existente, un objeto del mundo real, mediante los índices para poder diferenciar el mundo real del mundo de ficción. El predicado por sí solo, como un icono que es, como un signo de cualidad y potencialidad, no puede realizar esa función indicadora hacia el mundo existente; como afirma Peirce "el mundo real no puede distinguirse del mundo de ficción por ninguna descripción" (CP 2.337). Los índices son los encargados de dirigir la atención hacia lo que hay, y por ello, tanto los nombres propios, como los pronombres personales o demostrativos, los tonos, las miradas o los gestos emitidos por el hablante causan que el oyente preste atención a aquello que el hablante le indica, y para Peirce todos ellos son "índices del mundo real" (CP 2.337). Pero incluso el mundo de ficción de la literatura y la poesía puede ser considerado real, en tanto que como creación o imaginación de alguien es un hecho real, que una vez fijado no puede ser cambiado (CP 5.151).

Peirce explica cómo los índices que son nombres propios acaban transformándose en símbolos. Ello tiene lugar a través de un proceso en el que, primeramente, el nombre propio es un índice genuino hasta convertirse, a través de usos sucesivos, en un icono del índice, y finalmente en símbolo, que es interpretado como un icono del índice del objeto referido por el nombre. En este análisis de los nombres propios, que funcionan en su calidad de iconos como una copia o reproducción del nombre propio que en su origen fue índice, y así hasta que la costumbre y el hábito lo transforman en

[1] "Un signo o representamen es un primero que está en una relación triádica genuina tal con un segundo, llamado su objeto, que es capaz de determinar un tercero, llamado su intepretante, para que asuma *la misma relación triádica* con su objeto que aquella en la que se encuentra él mismo respecto del mismo objeto. La relación triádica es *genuina*, es decir, sus tres miembros están ligados por ella de manera tal que no consiste en ningún complejo de relaciones diádicas." (CP 2.274)

símbolos, Peirce anticipa la idea de Kripke de la cadena de comunicación para la determinación del referente (Cfr. R. Hilpinen 1995: 283-287).

La aserción

En el análisis peirceano de la aserción se observa con claridad su dimensión pragmática, al entenderla como una acción externa que involucra a dos protagonistas, el hablante y el oyente, con respecto a una proposición que el hablante cree verdadera y que tiene como objetivo que el oyente crea esa misma proposición. Para ello, el hablante toma sobre sí la responsabilidad de la verdad de la proposición. Así pues, un acto de aserción supone que, una vez formulada una proposición, una persona realiza un acto que le lleva a sujetarse a los castigos de la ley social o la ley moral, en el caso de que la proposición no sea verdadera.

Dos aspectos relevantes en el en el análisis de Peirce del acto de la aserción son: a) La nítida distinción entre el acto de la aserción y el acto de aprehender el significado de una proposición. b) La estrecha vinculación existente entre la aserción y el juicio. Esta vinculación puede resumirse así: el juicio es para Peirce un tipo de aserción que uno se hace a sí mismo, mientras que la aserción es el acto de aseverar una proposición, asumiendo una responsabilidad formal con respecto a su verdad, con la pretensión de que afecte a otros.

Otro aspecto muy importante de la aserción es lo que hoy en día llamaríamos, siguiendo a Austin, su aspecto ilocucionario, relacionado con las consecuencias punibles que se seguirían en el caso de que la proposición aseverada no fuera verdadera, consecuencias asumidas por el hablante que hace la aserción con respecto al oyente. A diferencia de los efectos perlocutivos del juicio, las consecuencias de la aserción son *formales o convencionales* (Austin, 2004: 74-81), lo que significa que forman parte de la fuerza ilocucionaria de la aserción. Al tener un carácter formal o convencional estas consecuencias son equivalentes a las que se derivarían de un contrato hecho ante notario, y por ello no se pueden entender como meros inconvenientes de tipo práctico para el hablante como en sucede con el juicio. En el acto de la aserción el carácter compromisorio sería mucho más acusado que en el juicio, justamente por ese efecto formal. En la aserción el hablante asume un compromiso y una responsabilidad hacia la verdad de lo aseverado en un grado tan formal que su incumplimiento lleva consigo el efecto convencional –previsto y regulado– de un castigo que el hablante conoce de antemano y acepta. Este tipo de situación es la propia de un acto desafortunado, que incurriría en el tipo de infortunio, que Austin denominó "abusos" (Austin, 2004: 60) y que es característico de los actos insinceros, ya que, como afirma Austin, "la palabra empeñada nos obliga" (Austin 2004: 55), y la aserción compromete a quien la realiza a creer o estar convencido de la verdad de la proposición.

El juicio

En el *juicio* Peirce distingue, por un lado, el aspecto lógico, centrado en la proposición, que es la materia sobre la que se ejerce el juicio y que es un tipo de signo, cuya naturaleza debe descifrar el lógico; y, por otro, el aspecto psicológico,

relativo a la naturaleza del acto de juzgar, y del que el lógico no debería preocuparse. Para aclarar la diferencia entre ese rasgo típicamente lógico del juicio y el rasgo psicológico de aceptar la proposición Peirce incorpora al análisis del juicio una nueva noción, la de valoración [*assent*]. El juicio es, pues, un acto de valoración, de aceptación de una proposición o una creencia, que puede ser de carácter subjetivo o de carácter público, identificándose este último con la aserción. De esta forma, el juicio pasa a ser definido como la aserción privada de una proposición, "la aserción a uno mismo" (CP 5.29), o dicho de otro modo, un acto interno valorativo con respecto a una proposición. Sería algo así como una aserción defectiva, porque la aserción como acto externo requiere de un hablante y un oyente, mientras que en el juicio el hablante y el oyente coinciden en la misma persona, destacándose de este modo la dimensión psicológica del juicio frente a la dimensión social y pública de la aserción.

La dimensión pragmatista de las actitudes cognitivas

La *creencia* es la actitud cognitiva en la que de manera más notoria se ponen de relieve los aspectos pragmatistas que caracterizan el pensamiento de Peirce, aunque en el análisis de la aserción y el juicio el elemento pragmatista del rechazo al dualismo hecho/valor ya estaba presente, al incorporar estas actitudes valoraciones, compromisos y responsabilidades hacia una proposición.

En los primeros escritos de Peirce, *la creencia* aparece caracterizada de modo similar al juicio, en el sentido de ser una conexión de ideas. De esta forma se vislumbra una estrecha relación entre la creencia y el juicio. Sin embargo, a pesar de esta coincidencia inicial, la creencia tiene una proyección que va más allá del juicio: la creencia es una conexión *habitual* de ideas, mientras que el juicio, en los primeros escritos peirceanos, es meramente una asociación de ideas. Más importante aún, la creencia está dirigida a la acción y proyectada por lo tanto al futuro, lo cual será su rasgo distintivo y definitorio, y que Peirce desarrolla en el contexto de su doctrina de la duda-creencia, propio de su pragmatismo.

Peirce indica tres propiedades de la creencia: primero, es algo de lo que somos conscientes; segundo, calma la irritación de la duda; y tercero, implica el establecimiento de un hábito en nuestra naturaleza, es decir, de una regla de acción; esta última propiedad es para Peirce la esencia de la creencia. Como regla de acción o hábito la creencia posee un carácter de generalidad y proyección hacia el futuro, constituyendo, entonces, un claro ejemplo de terceridad.

En consonancia con la creciente importancia que adquiere en su pragmaticismo la generalidad y su formulación a través de condicionales contrafácticos, los "would-bes", Peirce presenta nuevos matices en su consideración de la creencia: la creencia es creencia en una proposición de la que se está satisfecho y que tiene la forma de un hábito o ley general, que no sólo es activa en relación a una conducta presente o futura, sino que es activa también con respecto a una conducta pensada o imaginada, al considerar circunstancias posibles en las que la creencia daría forma a una acción, y que, si se produjese, determinaría efectivamente una acción concreta.

Referencias bibliográficas

Austin, J. L. (2004), *Como hacer cosas con palabras*. Tr. G. R. Carrió y E. A. Rabossi, Barcelona, Paidós.

Blanco Salgueiro, A. (2004), *Palabras al viento. Ensayo sobre la fuerza ilocucionaria*, Madrid, Trotta.

Hilpinen, R. (1995), 'Peirce on Language and reference', *Peirce and Contemporary Thought*, K. Laine Ketner (ed.), New York, Fordham University Press.

Peirce, C. S. (1931-1958), *Collected Papers* (CP), Cambridge (Mass.), Harvard University Press.

Tuzet, G. (2006), 'Responsible for Truth? Peirce on Judgment and Assertion', Pamplona, Grupo de Estudios Peirceanos.

El carácter fenoménico e intencional de los estados de ánimo y las emociones

Alberto Rubio Frutos
Universidad Autónoma de Madrid
Alberto.rubio@uam.es

Introducción

Piense en la última vez que ha experimentado una emoción. Con bastante probabilidad, pensará en un determinado escenario, en un determinado suceso que sea el causante y protagonista de esa emoción. Normalmente, habrá un objeto involucrado en dicha emoción, que quedará asociado a la misma, en función de la intensidad con la que se haya experimentado. Las emociones que se nos pasarán por la cabeza son desde la vergüenza hasta la alegría, desde el miedo hasta la ira, pasando por la envidia, los celos o la piedad.

Estos escenarios son los paradigmáticos del estudio de las emociones. Son casos en los que las emociones se experimentan intensamente, donde quedan claras nuestras respuestas corporales, nuestros pensamientos, un entorno específico y un objeto o un conjunto de objetos que son los claros protagonistas de esa emoción. La utilización de estos casos como paradigmáticos nos vienen dados por diversos motivos, en primer lugar, porque la concepción no académica de las vivencias emocionales se enmarca precisamente en escenarios de gran intensidad emocional; en segundo lugar, porque las discusiones en Filosofía de las Emociones se han centrado en cuál de esos elementos es más preponderante en los procesas emocionales. Por estos motivos, tanto en los experimentos realizados en ciencias cognitivas como los ejemplos a los que hacen alusión los filósofos tienen estas características.

En muchos casos, por ejemplo, se discute la naturaleza del objeto que nos provoca la emoción. Se discute si simplemente es la causa de la emoción, si es constitutivo de la misma, si es un objeto material, o si podemos hablar de las fantasías y recuerdos como objetos propiamente emocionales. Con este propósito, los escenarios empleados son intensos, para mostrar los argumentos de una parte y de otra con más intensidad.

Sin embargo, también se ha argumentado que la naturaleza de los estados emocionales es una parte fundamental de nuestra consciencia [Damasio (1999) p.p. 30; Ratcliffe (en prensa)]. Entre estas dos perspectivas antagónicas, se abre la posibilidad de analizar el verdadero alcance de las emociones en relación al resto de procesos mentales. Si las emociones no responden simplemente a situaciones de alerta y están involucradas en gran parte de las actividades que llevamos a cabo cotidianamente, queda por descubrir cuál es el verdadero alcance de las mismas.

El debate filosófico se centra en este aspecto en las relaciones entre emociones e intencionalidad. Desde la perspectiva de los escenarios centrados en situaciones de alerta y emociones intensas, las relaciones intencionales quedan establecidas en todos los casos, independientemente del diferente tipo de intencionalidad que se interprete, en función también del tipo de teoría de la emoción se suscriba. La mayoría de los escenarios descritos en estos casos se describen de hecho por una relación intencional que es emocional. Por ejemplo, en el caso del miedo, que es uno de los ejemplos más utilizados, las relaciones intencionales se describen entre el sujeto del miedo y el objeto causante del mismo. Del mismo modo ocurre con la pena, la alegría, el resentimiento, etc. Desde una perspectiva que sostiene que cierto tipo de procesamiento emocional está siempre involucrado en nuestros procesos mentales de carácter personal, la perspectiva es la contraria, toda intencionalidad es en parte emocional. Este último tipo de argumentación se sostiene desde diferentes posiciones y por muy diversas razones, pero por dar un ejemplo de carácter general, lo que se defiende es que no existe una relación neutra con el mundo, sino que está mediada en todos los casos por aspectos emocionales. Esta posición es difícil de sostener desde perspectivas cognitivistas de la emoción, pero es coherente con la idea de que la percepción del mundo no involucra sólo un acceso al conocimiento del mundo físico sino también a un mundo cargado de valores.

Estas dos perspectivas antagónicas de las relaciones entre intencionalidad y emoción, se sostienen también sobre la base de la inseparabilidad de las emociones de la intencionalidad. Mi propuesta es la de romper dicha relación para ofrecer un nuevo marco de estudio de las emociones que no esté determinado ni por escenarios excepcionales en nuestra vida cotidiana, ni por la difícil asunción de que las emociones están involucradas en todos y cada uno de nuestros procesos mentales.

La diferencia entre los estados de ánimo y las emociones

No pretendo, sin embargo, discutir las amplias ventajas que ofrece el estudio de emociones que están enmarcadas en un determinado contexto en el que podemos determinar más fácilmente las causas y las consecuencias de las mismas, así como a qué van dirigidas, su función en ese contexto, etc. Sino más bien lo que sostengo es que dichas aproximaciones dejan de lado procesos emocionales no enmarcados en episodios ni en situaciones determinadas, que han incluso contribuido a discutir su propia caracterización como una clase natural [Griffiths (1997), p.p. 173]

Una solución posible es la de considerar los estados de ánimo como un tipo especial de emociones cuya función es la de modular procesos emocionales de naturaleza episódica [Prinz (2004), p.p. 182-188]. El problema que presenta esta alternativa es que mantiene la idea de que las emociones tienen un carácter episódico que las distingue de los estados de ánimo. Este carácter episódico se traduce en dos criterios: la mayor duración y la distinta naturaleza de las causas y las consecuencias de los estados de ánimo. Estos dos criterios, junto al necesario carácter intencional de las emociones, son los han sido planteados principalmente

en la distinción entre estados de ánimo y emociones [Beedie, Terry and Lane, (2005) p.p. 847 – 878].

El primer criterio hace referencia a que las causas y las consecuencias de las emociones son directas e inmediatas. Sin embargo este hecho no es tan claro, porque alguien puede sentirse muy triste durante mucho tiempo después de que haya sucedido algo en su vida. La causa es inmediata pero las consecuencias no lo son. Por otro lado, las consecuencias pueden ser del mismo modo inmediato en un estado de ánimo, por ejemplo, podemos en un estado de ánimo de enfado y que esto tenga consecuencias directas e inmediatas, como es el hecho de enfadar a su vez a los que están a su alrededor.

El criterio de las causas y consecuencias tiene realmente su razón de ser en el de la duración que no es menos problemático. Puede haber emociones que tengan mayor duración que un estado de ánimo. No encuentro razón alguna para que nuestro estado de ánimo dure sólo desde que nos despertamos hasta que tomamos el primer sorbo de café, mientras que las emociones pueden durar mucho tiempo, como es el caso del amor, que siempre se ha defendido como una emoción y no como un estado de ánimo.

El carácter episódico de las emociones es por lo tanto discutible, y también lo es establecer una diferencia entre estados de ánimo y emociones basado en el mismo. Es el carácter intencional de toda emoción el criterio más contundente sobre el que basar esa distinción, y sobre el que justificar a su vez una aproximación al estudio de las emociones constituida sobre su contenido intencional. El carácter episódico de los casos sobre el que nos basamos para el estudio de las emociones viene dado no por el necesario carácter episódico de las emociones en sí mismas, sino más bien como consecuencia de que toda emoción es intencional, y que resulta más sencillo capturar dicha intencionalidad en el análisis de situaciones y episodios emocionales específicos.

La intencionalidad de las emociones

Los estados de ánimo vienen determinados más bien por aspectos fenoménicos, es decir, se definen más bien en virtud de lo que sentimos y no en virtud de la relación intencional que guardamos con un determinado aspecto del mundo como, por ejemplo, cuando nos sentimos irritados alguna mañana. Por el contrario, en el caso del amor o del odio, lo que hace de éste una emoción es que está dirigido hacia alguien o algo, y su carácter fenoménico está asociado constitutivamente a su contenido intencional, en este caso, el objeto de su amor o de su odio. Esta posición ha sido defendida explícitamente por varios autores:

1. Ronald De Sousa ha defendido que las emociones deben ser explicadas a través de mecanismos que incluyan acciones intencionales, porque su función es la de motivar la acción humana. Al haber objetos que forman parte de las emociones, éstas pueden definirse en virtud de su contenido intencional [De Sousa (1987)].
2. Goldie defiende que las emociones son también intencionales: “los pensamientos y los sentimientos involucrados en las emociones tienen

una *dirección hacia* un objeto". [Goldie (2000), p.p 19] Las emociones son *sentimientos hacia cosas*. En este caso, el contenido intencional de las emociones viene dado por su carácter fenoménico.

3. Gunther ha defendido que la distinción *fregeana* entre fuerza y contenido no es respetada en el caso de las emociones. Éstas no serán como las actitudes proposicionales porque la fuerza no es independiente del contenido. En las actitudes proposicionales podemos distinguir entre qué decimos y cómo lo decimos. La diferencia con el contenido emocional es que la actitud está incluida necesariamente en lo que se está representando emocionalmente. Es más, cualquier expresión con contenido emocional no es distinguible de expresar nuestros sentimientos, nuestra actitud hacia algo. Es decir, que cualquier tipo de contenido emocional tiene una actitud y un sentimiento asociado a él [Gunter (2004), p.p. 43-55].

Todos estos argumentos provienen de interpretaciones muy diferentes de la naturaleza de las emociones. Sin embargo, no sostienen necesariamente un modelo episódico de corta duración de las emociones. Lo que sí se infiere de sus argumentos es que los estados de ánimo son diferentes a las emociones. Si por las razones que hemos argumentado en la anterior sección, mantenemos que realizar una distinción profunda entre ambos fenómenos es poco plausible, debemos llegar a las siguientes conclusiones:

1- Las emociones en tanto que estados de ánimo pueden carecer de objetos intencionales, y poseer exclusivamente aspectos fenoménicos, que no motiven conducta alguna.

2- La direccionalidad de las emociones también es discutible en tanto que hay emociones que poseen cierto carácter disposicional, pero dichas disposiciones pueden perfectamente carecer de una direccionalidad hacia un objeto.

3- Es posible por tanto concebir una emoción sin contenido mental, entendiendo por ello un tipo de representación o acción que sea el propio contenido de la emoción. Los estados de ánimo son en ocasiones emociones "puras", en el sentido de que están constituidas por sentimientos que no son traducibles en contenidos ni conceptuales ni no conceptuales.

Conclusión

Durante esta presentación nos hemos centrado en la crítica a una determinada aproximación a los escenarios tradicionales utilizados en el estudio de las emociones. Dicha aproximación involucra dos aspectos: que los estados de ánimo no son emociones genuinas, y que estas últimas tienen necesariamente contenido intencional. Mediante el cuestionamiento de estas posiciones hemos abierto la posibilidad de estudiar fenómenos poco explorados anteriormente.

Referencias bibliográficas

Beedie, T. y L. (2005), 'Distinctions Between Emotion and Mood', *Cognition and Emotion* 19, pp. 847-78.

Damasio, A. (2000), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York, Harcourt Brace.

Griffiths, P. (1997), *What Emotions Really Are: The Problem of Psychological Categories,* Chicago, University of Chicago Press.

Goldie, P. (2000), *The Emotions: A Philosophical Exploration,* Oxford, Oxford University Press.

Gunther Y. H. (2004), 'The Phenomenology and Intentionality of Emotions', *Philosophical Studies* 117, pp. 43-55.

Prinz, J. (2004), *Gut Reactions: a Perceptual Theory of Emotion*, New York, Oxford University Press.

Ratcliffe, M. (en prensa), 'The Phenomenology of Mood and the Meaning of Life', en Goldie, P. (ed), *Oxford Handbook of Philosophy of Emotion*, Oxford, Oxford University Press.

Nihilism, indifference, and ontological commitment

Pablo Rychter
Universitat de Girona, LOGOS/ Somerville College, Oxford
prychterus@yahoo.com

Introduction

Proponents of austere ontologies generally face the problem of explaining the fact that ordinary speakers (which may be themselves in their unphilosophical moments) seem to refer to, and quantify over, those very entities that, by the austere ontologist's lights, do not really exist. To give just some examples, nominalists have to explain why we seem to be saying something true by uttering 'the number of planets is 9', presentists have to explain our apparent reference to past objects, and nihilists about composition have to explain our apparent reference to composite things. How are we to understand our talk apparently about Fs, if Fs do not really exist? Let us call this the *reconciliation problem*, the problem of explaining how ordinary, apparently committal talk can be reconciled with austere ontological doctrines.

I will focus here on how the *reconciliation problem* arises for mereological nihilism –the view according to which there are no composite objects, no things that have other things as proper parts. As the view is commonly discussed, it implies that there are no chairs, animals, persons, etc. It may thus seem that the view contradicts common sense and ordinary thought. But according to van Inwagen, this is not really so: mereological nihilism "radical though it is, does not contradict our ordinary beliefs". [van Inwagen (1990), p. 98].[1] To some important extent, I agree with van Inwagen that there is no contradiction, but I am dissatisfied with his explanation of why this is so. My main purpose here is to offer a more complete and satisfying explanation on the nihilist's behalf.

Van Inwagen and ontological commitment

Let us see first why van Inwagen's explanation is unsatisfying, focusing our attention on a particular example. Suppose that in an ordinary situation, you utter the following sentence:

(1) There is a chair in the room next door.

Your utterance of (1) seems to be true in the envisaged situation, and it also seems to carry commitment to chairs –i.e., it seems that chairs must exist in order for the

[1] Van Inwagen makes this remark about his own view on ontology, which is not exactly mereological nihilism but something close to it. Since the difference is irrelevant for our purposes, I will put it aside in what follows, and talk as if van Inwagen were a nihilist proper.

utterance to be true. But if mereological nihilism is true, at least one of these two appearances must be wrong: either your utterance is not really true, or it is not really committed to chairs. Van Inwagen clearly rejects the first alternative. As he makes the point:

> My position, therefore, is that when people say things in the ordinary business of life by uttering sentences that start "There are chairs…" or "There are stars…", they very often say things that are literally true. [van Inwagen (1990), p. 102].

So his view must be that, contrary to appearances, your utterance of (1) does not carry commitment to chairs. How can this be? What explains the appearances to the contrary? What does your utterance mean if not something that requires the existence of chairs for it to be true? This is where van Inwagen leaves the job unfinished. What he actually offers at this point is an indication of how an apparently committal sentences can be paraphrased into sentences that do not even appear to carry commitment to composites. So (1), for instance, could be paraphrased as follows:

(2) There are simples arranged chairwise in my office.

But van Inwagen does not offer an explanation of how these paraphrases are to be understood, i.e. of what exactly the relation between the original and the paraphrase is supposed to be. And this is precisely why his solution to the reconciliation problem is unsatisfying, or at best incomplete. Notice that a prominent view about the nature of paraphrase is not available to van Inwagen. He cannot say that the paraphrase is a proposed true replacement for an original that is close to truth but nevertheless false. He cannot say this because he has admitted that the original is already "literally true". Rather, van Inwagen must be presenting the paraphrases in a "hermeneutic spirtit", as elucidations of what the originals really mean. But, in general, the idea that paraphrases capture the meaning of the originals is puzzling: in what sense are paraphrases an improvement over the originals if they have exactly the same meaning? The idea that (2) captures the meaning of (1) is not advisable for van Inwagen. And in fact, he explicitly says that "paraphrases are not supposed to capture the meaning of their originals" (112). But then it is unclear what his view about this issue finally is, and thus it is in the end not clear why nihilism "does not contradict our ordinary beliefs".

As I will argue in the next section, I think a nihilist like van Inwagen may well drop the project of paraphrase altogether. Paraphrases do not play any relevant role in the explanation I favor of why there is (to some extent) no contradiction between nihilism and ordinary belief. However, it is worth noticing here a possible view that is still open to van Inwagen: (2) does not capture the meaning of (1) but still captures the *truth conditions* of your *particular utterance* of (1). This idea can be accommodated in the picture offered below.

To sum up: van Inwagen's view is that your utterance of (1) is literally true. And this implies, it seems, that your utterance does not carry commitment to chairs. But we do not have a satisfactory explanation of why this is so. Appearances to the contrary are not explained. In this respect, van Inwagen's

position is almost as unsatisfactory as the position of an error theorist who said that (1) is not true without bothering to explain why it seems that it is.

An alternative approach

Here is a hypothesis that, if true, would explain why your utterance of (1) is not committed to chairs and why it *seems* to be so: the context on which your utterance has its "essential effect" is (unlike the philosophy room) such that the existence of chairs is presupposed. This is a false presupposition, according to the nihilist, but one that ordinary speakers nevertheless make. Because the existence of chairs is already presupposed in the context, the point of your utterance is *not* to assert –among other things- that chairs exist. Rather, the content that you assert, the point you intend to make by uttering (1), is *indifferent* as to whether chairs exist –it neither requires that chairs exist, nor does it require that chairs do not exist. But although the existence of chairs is not what you want to assert, presupposing that chairs exist is a useful means for making your point. Not making this presupposition may imply going through a detour that would not serve well your communicative purposes. Of course, in other occasions (1) could have been used to assert something that does imply the existence of chairs. For instance, in the context of a philosophical discussion, one could use (1) as a premise in an argument against mereological nihilism. In that case, the primarily asserted content–what the speaker primarily wants to communicate- would not be indifferent with respect to whether chairs exist. But in an ordinary context like the one in our example, the content of your assertion is consistent with mereological nihilism.

In this respect, your utterance is similar to what allegedly happens in the following case:

A: The man drinking martini is a philosopher.

B: He is not really drinking martini.

A: Whatever. My point was that he is a philosopher.

Character A does not take himself to be accountable for the apparent consequence of his assertion. His response to B shows the characteristic *impatience* that some have identified as a common reaction to the ontological scruples [cf. Yablo (2000), Eklund (2005)]. One is impatient about the observation that *not p* when *p* is not part of what one said –when one is indifferent or even in disagreement about *p*. It seems, then, that you are entitled to be indifferent about the truth value of your presuppositions –that you are not required to presuppose only what you think is true. A presupposition is an instrument that helps you get to the asserted content, and it can perform this function perfectly well even if you do not think it true, and even if it is actually false. So, in presupposing a proposition that entails the falsity of mereological nihilism, you may still be perfectly neutral about whether this doctrine is true.

The main point so far is this: if it is true that your utterance of (1) only presupposes and does not assert that chairs exist, we have an explanation of why

nihilism does not contradict ordinary assertions. The explanation has the virtue of also making clear why your utterance *seems* to be committed to chairs even if it in fact isn't. Notice that in this solution to the reconciliation problem, paraphrases do not play any important role. Even assuming that (2) captures the truth conditions of your utterance of (1), the existence of the paraphrase itself is irrelevant, and not what plays the "legitimazing role" attributed to it. Your utterance of (1) is acceptable not because it is paraphrasable into (2). Rather, it is acceptable because it expresses a true content –whether or not that content can also be captured by a suitable paraphrasis.

Let me say something about how the present proposal relates to others in the recent literature. That speakers are partially indifferent about the ontological commitments of some of their utterances is stressed by Yablo (2000), Eklund (2005), and it is arguably also the view of van Inwagen. Yablo focus on cases of indifference about the existence of abstracta, and goes on to explain this indifference as a result of the speaker's engagement in some kind of pretense: speakers are indifferent about the implication *p* of what they literally say because they are only *pretending* that things are such that *p* is the case. What is characteristic of Eklund's "indifferentism", on the other hand, is the rejection of this explanation of indifference in terms of pretense: the "don't care attitude" that speakers sometimes adopt toward some aspects of what they say need not be the result of engaging in pretense. The hypothesis that I am putting forward –that the existence of chairs is a presupposed but not asserted by your utterance of (1)— may be seen as an elaboration of Eklunds's position, but one according to which the gap between fictionalism and indifferentism is not so wide. When speakers assume a "don't care" attitude, they are still aiming at some sort of truth –"truth under false pressupositions", so to speak. Also, the hypothesis makes explicit why the "don't care" attitude is not gratuitous or irresponsible. The attitude is justified because it is the price that speakers pay for an efficient way of expressing the content that they do care about. This is just a case of the general fact that false presuppositions are often useful tools to express certain contents.

Before closing, let me make a couple of clarifications and anticipate some objections. First, it is important to appreciate how limited the proposed reconciliation is. For all I have said, it may well be that ordinary speakers *believe* that chairs exist, and that they are consequently disposed to answer positively to the question 'do chairs exist?'. But van Inwagen's original point, and my elaboration of it, is not that ordinary speakers are convinced nihilists, or agnostic about the issue of nihilism vs anti-nihilism. The point was simply that they are not constantly contradicting nihilism with each of their every-day utterances of sentences like (1). This is consistent with the likely possibility that, were the question brought out, they would assert the existence of chairs and plainly reject nihilism.

Second, it must be emphasized that the fact that the existence of chairs is presupposed but not asserted in ordinary utterances of (1) is a hypothesis that stands or falls independently of nihilism. The hypothesis concerns language users, not ontology. If it is true, it is good news for a mereological nihilist like van

Inwagen. But the reasons for believing the hypothesis will still hold were mereological nihilism false. (Notice that this point also applies to the heremeutic fictionalism advocated by Yablo).

References

Eklund, M. (2005), 'Fiction, Indifference, and Ontology', *Philosophy and Phenomenological Research* LXXI, pp. 557-59.

Van Inwagen, P. (1990), *Material Beings*, Ithaca, Cornell University Press.

Yablo, S. (2000), 'Apriority and Existence', in Bogossian, P. and Peacocke C. (eds.), *New Essays on the A Priori*, Oxford, Oxford University Press, pp. 197-228.

Roots to numerical cognition

Mario Santos-Sousa
Universidad Autónoma de Madrid
msansou@gmail.com

Current cognitive studies suggest a twofold distinction among infant's innate numerical capacities: a sense of small cardinal size and a rough sense of large cardinal size, which allow us, at best, to discriminate small collections of objects and approximate larger numerical quantities. In view of their obvious limitations, this paper examines other cognitive mechanisms – crucially, ones that are not specific to number – that might support the acquisition of more sophisticated numerical skills by providing the relevant recursive structure. I will discuss how the capacity for performing recursive computations may be implemented by different systems (linguistic, sensory-motor, and visual) and evaluate the extent to which these mechanisms conform to the principles that govern arithmetic reasoning and thus represent a reliable way of acquiring arithmetical knowledge. On the view which I recommend, numerical competence lies at the interplay of different cognitive capacities, which are not specific to number but seem to have their cognitive homes elsewhere, that are appropriately co-opted as an individual's abilities in mathematics grow.

Introduction

In recent years, the study of our basic mathematical capacities has received much attention within cognitive science. This research has largely centered on our facility with numbers—under the working assumption that having a good grip on the mechanisms of basic numerical thinking may help us to tackle more complex cases. As a result, substantial evidence on our primitive numerical capacities has been gathered, thus giving us valuable insight into how (and why) the human mind is suited for mathematics. Current studies suggest that these capacities have two basic components: (a) a number-specific, analog system of representation known as the *accumulator*, which accounts for our ability to respond differentially to approximate number, and (b) a mechanism of object-based attention that allows us to track a small number of objects by means of a comparably small number of representations (object files) primarily based on the spatio-temporal properties of those objects. For a comprehensive review, see [Feigenson et. al. (2004)].

These capacities are nevertheless embarrassingly limited when it comes to explaining the development of more sophisticated numerical abilities, such as our mastery of the sequence of natural numbers. On the one hand, the accumulator represents numerical properties only approximately, using a system of continuous mental magnitudes. The concept of a natural number, however, includes the notion that each such number is discrete and has a unique successor. But there is nothing about magnitudes that enforces this idea, since magnitudes are continuous and do

not have unique successors. On the other, the object-indexing system is confined to the detection of small collections of objects within the span of immediate apprehension (that is, up to four) and thus cannot support the notion of infinity. It fails to ground the idea that it is possible to add one indefinitely. How, then, do these primitive capacities give place to more sophisticated ones? A common thought is that they need to be supplemented by a capacity for carrying out recursive computations (which would allow us to entertain the idea that the natural numbers form an infinite ordered sequence). What this capacity exactly amounts to has been a matter of some dispute. According to some proposals, it depends crucially on the recursivity of language. Others trace them back to our sensory-motor skills or even to our visual means of identifying structural patterns. I will examine them in turn.

Language

Language may play an important, perhaps an essential role in the acquisition of the natural number sequence, by bestowing on it the property of "discrete infinity". I will analyze its possible contribution from two different perspectives: (a) along Chomskian lines, with regard to the language faculty and its core mechanism for recursion, and (b) in connection with natural language, following Carey's "bootstrapping" account.

Chomsky once suggested that our understanding of number developed as an evolutionary by-product of the human language faculty: "We might think of the human number faculty as essentially an 'abstraction' from human language, preserving the mechanism of discrete infinity and eliminating the other special features of language" [Chomsky (1988), p. 169]. In more recent work, Hauser, Chomsky and Fitch [2002] approach the relation between language and number slightly differently. Their idea is that both share the property of discrete infinity because both systems spring from an underlying capacity to perform recursive computations.

Following Bloom [1994], I believe that these suggestions are subject to a criticism based on Chomsky's own observation that the possession of discrete infinity confers no selective advantage in the domain of numerical cognition. For the same lack of selective pressure makes it unlikely that numerical cognition could inherit the generative property of language or of some other underlying recursive capacity.

An alternative view is that children acquire the concept of natural (or counting) number by "bootstrapping" their way through the count sequence. Children's sensitivity to linguistic cues may guide them to an initial stage of number word understanding, while the rest of their knowledge emerges through an understanding of counting. As Carey succinctly puts it: "The count list ('one, two, three ...') is a system of representation that has the power to represent the positive integers, so long as it contains a generative system for creating an infinite list. When deployed in counting, it provides a representation of exact integer values based on the successor function. That is, when applied in order, in one-one

correspondence with the individuals in a set, the ordinal position of the last number word in the count provides a representation of the cardinal value of the set—of how many individuals it contains" [Carey (2004), p. 63].

There is, however, an important objection to the role of language in accounting for numerical competence. Cases of spared calculation and number comprehension abilities have been described in patients with severe impairment of semantic processing (called semantic dementia), as well as conditions of spared semantic processing but impaired calculation (known as dyscalculia). Thus, a double dissociation obtains between our linguistic and our numerical capacities [Butterworth (1999)]. In view of this, language can have, at most, a facilitating role but seems to be neither necessary nor sufficient for our grasp of number.

Sensory-motor activities

Perhaps children acquire natural number concepts by mapping them from groupings of physical objects. The claim is that the general properties of the natural number sequence are modeled on everyday experience with concrete objects. "Children come to learn the meanings of 'set', 'number', 'addition' and to accept basic truths of arithmetic by engaging in *activities* of collecting and segregating" [Kitcher (1984), pp. 107-108]. A recent reappraisal of this (Millian) line of thought builds on the notion of *conceptual metaphors* "[whose] primary function is to allow us to reason about relatively abstract domains using the inferential structure of relatively concrete domains" [Lakoff and Núñez (2000), p. 42].

Lakoff and Núñez introduce the "Arithmetic is Object Collection" metaphor to link the domain of physical objects to the domain of numbers, capturing the idea of discreteness and order. The underlying thought is that numbers inherit those properties through a mapping from the source domain (collections of concrete physical objects and operations performed on them) to the target domain (sets of numbers and operations defined over them).

A further metaphor, "The Basic Metaphor of Infinity", projects the notion of an infinite set or sequence—generated by repeating a particular mathematical operation—from experience with repeated operations on collections of physical objects. Grasp of the successor function would rests on the child's experience that it is always possible "in principle" to add another item to a given collection of objects.

However, given our interaction with a limited number of objects, why should we feel entitled to draw these general conclusions (e.g. that it is *always* possible "in principle" to add another object to a given collection of objects)? Everyday experience with physical objects, which provides the source domain for the metaphors, doesn't always exhibit the properties that these metaphors are supposed to supply. And yet, it seems that experience *does* trigger *general* belief forming dispositions about objects: that they are discrete, that they can be collected into larger and larger collections, or put (ordered) into different arrays. Lakoff and Nuñez's point is that "such everyday experiences…result in connections between sensory-motor physical operations…and arithmetic operations…[which] *constitute a*

conceptual metaphor [that] is learned at an early age, prior to any formal arithmetic training. Indeed, arithmetic training assumes this unconscious conceptual (non linguistic!) metaphor" [(2000), pp. 54-55].

Visualization

In addition to mapping mathematical properties from experience with physical objects, we can have a grasp of some mathematical structures by means of visual representations. In a recent book, Marcus Giaquinto [(2007), ch. 6 & 11] tries to show that some infinite structures can be known by visual means. Indeed, there is a large body of empirical evidence suggesting that aspects of numerical representations are visuo-spatial in nature [Dehaene et. al. (2008)].

According this proposal, we can have a grasp of the natural numbers as forming an ordered (infinite) sequence by visualizing them in a spatial array. But how? We could achieve this kind of understanding by reading off the relevant properties from an actual drawing. However, how do we know what is "relevant" and what is not? There is a threat of taking for granted what we try to explain: numbers may be amenable to this type of representation *because* they form an ordered set. It is the ordinal nature of numbers which gives rise to a visuo-spatial format of representation, not the other way round.

A second possibility is that number sense representations triggered by the accumulator and the object-indexing system are associated with positions in egocentric space and with positions on a visual line. This seems to occur spontaneously and without conscious effort. Consider our grasp of the natural number sequence as a set of evenly spaced vertical marks on a horizontal line, with a single leftmost initial mark, continuing endlessly to the right such that every mark, no matter how far to the right (since there is no terminal mark), is ultimately reachable. One mark precedes another if it is left to it and succeeds another if it is right to it (left-right ordering). To be sure, this does not yield an infinitely extended visual image when you try to visualize it. That's impossible, of course. But a description as the one just given recursively specifies a line with no right end. Being able to visualize this feature amounts to having certain dispositions which tap the nature of the structure they represent, a well-ordered sequence with a single initial element and no terminal element: the natural number sequence.

Conclusion

This paper puts forward a view of numerical cognition as lying at the interplay of different cognitive capacities—not all of them specifically "numerical". This approach might prove useful in accounting for our distinctive competence in other mathematical domains, with various mechanisms being co-opted as an individual's abilities in mathematics grow.

References

Bloom, P. (1994), 'Generativity within language and other cognitive domains', *Cognition* 51, pp. 177-89.

Butterworth, B. (1999), *The Mathematical Brain*, London, Macmillan.

Carey, S. (2004), 'Bootstrapping & the origin of concepts', *Daedalus*, Winter 2004, pp. 59-68.

Chomsky, N. (1988), *Language and problems of knowledge*, Cambridge, MA, MIT Press.

Dehaene, S., Izard, V., Spelke, E. and Pica, P. (2008), 'Log or Linear? Distinct intuitions of the number scale in Western and Amazonian Indigene Cultures', *Science* 320.

Feigenson, L., Dehaene, S. and Spelke, E. (2004), 'Core systems of number', *Trends in Cognitive Sciences* 8, n. 7, pp. 307-14.

Giaquinto, M. (2007), *Visual Thinking in Mathematics. An Epistemological Study,* Oxford, Oxford University Press.

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002), 'The faculty of language: What is it, who has it, and how did it evolve?', *Science* 298, pp. 1569-79.

Kitcher, P. (1987), *The Nature of Mathematical Knowledge*, New York/Oxford, Oxford University Press.

Lakoff, G. and Núñez, R. E. (2000), *Where mathematics comes from: How the embodied mind brings mathematics into being*, New York, Basic Books.

Does non-linguistic systematicity tell against mental representation in a lot?

Víctor Martín Verdejo Aparicio
Universitat Autònoma de Barcelona
Victor.Verdejo@uab.es

For many years now both philosophers and cognitive scientists have paid attention to the phenomenon of systematicity. While disagreeing in many other respects, the following seems to be common ground: systematicity phenomena demand the postulation of mental representation (MR henceforth) with associated specifiable constituent structure. Roughly, if a subject's capacity to think that Mary loves John nomologically entails her capacity to think that John loves Mary, that must be, according to the familiar Fodorian diagnosis (e.g., Fodor 1987, pp. 150-1; 1997, pp. 111-12; 2004, pp. 36-7; Fodor and Pylyshyn 1988, pp. 33-7; Fodor and McLaughlin 1990, pp. 200-3), because the MR involved in the subject's thinking that Mary loves John has the very same constituents than the MR involved in the subject's thinking that John loves Mary. However, does the postulation of structured MR involve the postulation of a system of neurophysiologically identifiable symbols –i.e. does structured MR involve a Language of Thought (LOT henceforth)? Looking for a negative answer to that question, some authors belonging to the connectionist landscape have paid attention to the features of *non*-linguistic systematicity, i.e., systematicity exhibited in the cognition of non-linguistic domains such as vision, audition, music, algebra, etc. More precisely, Cummins and allies (Cummins 1996; Cummins et al. 2001; Cummins et al. 2005) have strongly argued that, since MR in a LOT is MR that *shares* structure with the linguistic domain, and since systematicity is also a cognitive phenomenon to be found in domains other than language, either

> a) some MR is not LOT-MR –i.e., some MR shares non-linguistic structure–

or else

> b) some kind of encoding –i.e, MR that does not share structure with the domain– is needed in order for LOT-MR to account for non-linguistic systematicity.

That is an embarrassing dilemma for the LOT defender: a) is certainly self-defeating for LOT as a general model of cognition while b) involves the legitimacy of encodings –the kind of MR favoured by connectionist models– in explanations of systematicity, and for that matter, in explanations of linguistic systematicity. In short, if Cummins et al.'s considerations are sound, the classical argument from systematicity to MRs in a LOT is certainly blocked.

Cummins et al.'s argument receives indirect support, on the one hand, from the fact that it has been actually accepted by authors coming from the classicist side of

this debate. For instance, Davis has conceded that "if the argument from linguistic systematicity to classical representations is sound, then so is an argument from visual systematicity to non-classical representation" (Davis 2005, p. 403). On the other hand, and strikingly enough, Cummins et al.'s conclusion seems to accord well with the recently presented Fodorian view that iconic representation involves a distinctive kind of compositionality. In particular, according to Fodor (2007 and 2008, chap. 6), iconic representation, as opposed to discursive or linguistic representation, does not contain canonical constituents. As Fodor puts it, "[a]n icon is a homogeneous kind of symbol from both the syntactic and the semantic point of view" (Fodor 2008, p. 174). But then, it is easy to interpret Fodor's considerations as allowing for a kind of MR that is out of the scope of LOT. More precisely, if icons are structurally homogeneous symbols –that is, if there is no way of distinguishing canonical parts from mere parts in them– and if LOT-MR must share structure with language, then it pretty nearly follows that icons cannot belong to the representational possibilities within LOT, which would then squarely require heterogeneous symbols –that is, symbols with distinctive semantic and syntactic properties.[1]

That being so, I think, for reasons to follow, that Cummins et al.'s argument is fatally flawed. In my view it is clear, and so I will presently defend, that LOT-MR (may but) does not have to *share* (as opposed to *encode*) the structure of language, and hence, that the alleged need of encodings in explanations of *non*-linguistic systematicity is no embarrassment for LOT. If I am right, Cummins et al.'s particular systematicity argument against LOT, apparent concessions to the contrary notwithstanding, is not forthcoming.

We can immediately cast doubts on Cummins et al.'s strategy by just considering Fodor and Pylyshyn's original presentation of the pro-LOT systematicity argument in which they paused to remark that "linguistic capacity is a paradigm of systematic cognition, but it's wildly unlikely that it's the only example" (Fodor and Pylyshyn 1988, p. 37). Therefore, that there is non-linguistic systematicity can hardly be news for the LOT defender. Furthermore, a series of authors have emphasized that LOT-MR is equally appealing when considering non-linguistic domains and specifically perception (e.g., McLaughlin 1993; García-Carpintero 1995). For instance, to be able to perceive that the ball is red and that the square is blue must involve the capacity to perceive that the ball is blue and that the square is red, which would be explained by the postulation of the corresponding symbols in a LOT.

---

[1] In fact, Fodor's considerations in this respect are doubly puzzling. It is not only that the postulation of syntactically and semantically homogeneous symbols strongly suggests that iconic representation is not of the LOT kind. Moreover, it is hard to see, at least without further articulation, how the postulation of such icons would figure in explanations of, as it were, iconic systematicity, since systematicity phenomena require, not only the existence of some kind of compositionality whatsoever, but of compositionality that allows identifying the contribution of particular constituents in different arrangements of representations.

However, to show that Cummins et al.'s argument is wrong I will defend the following instance of modus ponens: 1) If encodings are a kind of LOT-MR in the first place, then there is no embarrassment for LOT from the use of encodings in explanations of systematicity (be it linguistic or otherwise). 2) Encodings are, to all intents and purposes, just the kind of MR to be expected within LOT. Conclusion: No embarrassment for LOT presently. The crucial premise is of course 2). As the authors point out, encodings –such as tensor-product networks (Smolensky et al. 1992) or Gödel numbering (Van Gelder 1990)– can preserve the structure responsible for systematicity phenomena without actually sharing that structure. In the linguistic case for instance, Gödel numbers can encode the structure of grammatical sentences via uniquely factorable numbers obtained through the product of a series of numbers n^m, where m are natural numbers associated to words and n prime numbers associated to the word's syntactic position. In my view, we can only reject encodings –such as Gödel numbering– as genuine LOT-MRs by requiring of LOT some of the following:

a) Explicit representation of semantic axioms and rules. If LOT requires MR with explicit representation of rules then it is true that encodings cannot be MRs within LOT. However it is a very old point that LOT does not require such a thing:

> RTM [the Representational Theory of the Mind or LOT] says that the contents of a sequence of attitudes that constitutes a mental process must be expressed by explicit tokenings of mental representations. But the rules that determine the course of the transformation of these representations –modus ponens, 'wh'-movement, 'get the queen out early', or whatever– need not themselves ever be explicit. They can be emergents out of explicitly represented procedures of implementation, or out of hardware structures, or both. Roughly: According to RTM, programs –corresponding to the 'laws of thought'– *may* be explicitly represented; but 'data structures'– corresponding to the contents of thoughts– *have to be* [Fodor (1987), p. 25, his emphasis].

b) Existence of particular spatial (like in written language) or temporal (like in spoken language) relations among constituents. If LOT demands that the language of thought be alike to natural languages regarding spatio-temporal relations among constituents, then it is pretty certain that encodings cannot be part of a LOT: e.g., the Gödel numbers '437.400' and of '1.093.500' do not even remotely exhibit the spatio-temporal relations among constituents found in the sentences that they can serve to encode –i.e., 'John loves Mary' and 'Mary loves John' respectively. It is nonetheless also settled from the outset, as García-Carpintero (1996) has already emphasized, that all that LOT requires is a functional relation among constituents:

> […] since the relation [of adjacency among symbolic expressions] to be physically realized is *functional* adjacency, there is no necessity that physical instantiations of adjacent symbols be *spatially* adjacent. Similarly, although complex expression are made out of atomic elements, and the distinction between atomic and complex symbols must somehow be physically instantiated, there is no necessity that a token of an atomic symbol be assigned a smaller region in space than a token of a complex symbol; even a

> token of a complex symbol of which it is a constituent. [Fodor and Pylyshyn (1988), p. 57, their emphasis].

c) Existence of a one-to-one correspondence between categories found at a high and at a low level of description. This point can be put in terms of Marr's celebrated levels of description (Marr 1982). According to Marr, even if there must be some explanatory relation between higher and lower levels of description of cognitive phenomena, it is only to be expected that the relation between (higher and lower) categories is one-to-many. Therefore, it is only to be expected too, that LOT MRs do not keep a one to one correspondence with high-level categories. The word 'Mary' can be implemented by '4' or by '25' in Gödel numbering, depending on whether it is 'Mary' in 'Mary loves John' or in 'John loves Mary' that is at stake. However, that fact does not prevent Gödel numbering of being a representational possibility within LOT itself.

On reflection, that encodings are a kind of LOT-MR can hardly be surprising. True, numbers or matrices are not structurally like sentences in a language. However, the distinctive nature of numbers and matrices *as encodings* in explanations of linguistic systematicity is that they can structurally *work as* sentences, hence as semantic/syntactic units, hence as semantic/syntactic units with identifiable physical properties in a LOT. If it is not by way of additional implausible requirements on LOT, there is little reason to think that encodings cannot be MR in a LOT. A fortiori, there is little reason to think that LOT is not able to accommodate, via encodings of the suitable sort, systematicity phenomena in domains other than language. It remains of course an open (ultimately empirical) question how exactly systematicity phenomena of the various non-linguistic sorts are accounted for within a LOT model. In the meantime, and until further evidence is brought to bear, it is worth noting that there doesn't seem to be much to the contention that non-linguistic systematicity tells against LOT.

References

Cummins, R. (1996), 'Systematicity', *The Journal of Philosophy* 93:12, pp. 591-614.

Cummins, R., Blackmon, J., Byrd, D., Poirier, P., Roth, M. and Schwarz, G. (2001), 'Systematicity and the Cognition of Structured Domains', *The Journal of Philosophy* 98:4, pp. 167-87.

Cummins, R., Blackmon, J., Byrd, D., Lee, A. and Roth, M. (2005), 'What Systematicity Isn't: Reply to Davis', *Journal of Philosophical Research* 30, pp. 405-8.

Davis, W. (2005), 'On Begging the Systematicity Question', *Journal of Philosophical Research* 30, pp. 399-404.

Fodor, J. A. (1987), *Psychosemantics*, Cambridge, Massachusetts, MIT Press.

— (2007), 'The Revenge of the Given', in McLaughlin, B. and Cohen, J. (eds.), *Contemporary Debates in Philosophy of Mind*, Oxford, Blackwell, pp. 105-16.

— (2008), *LOT2*, New York, Oxford University Press.

Fodor, J. A. and Pylyshyn, Z. (1988), 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* 28, pp. 3-71.

García-Carpintero, M. (1995), 'The Philosophical Import of Connectionism: A Critical Notice of Andy Clark's Associative Engines', *Mind & Language* 10:4, pp. 370-401.

— (1996), 'Two Spurious Varieties of Compositionality', *Minds & Machines* 6:2, pp. 159-172.

Marr, D. (1982), *Vision*, San Francisco, Freeman.

Mclaughlin, B. (1993), 'The Connectionism/Classicism Battle to Win Souls', *Philosophical Studies* 71, pp. 163-90.

Smolensky, P., Legendre, G. and Miyata, Y. (1992), 'Principles for an Integrated Connectionist/Symbolic Theory of Higher Cognition', Institute of Cognitive Science, University of Colorado, Technical Report 92-08.

Van Gelder, T. (1990), 'Compositionality: A Connectionist Variation on a Classical Theme', *Cognitive Science* 14, pp. 355-84.

El contenido y el problema de la creencia*

Ignacio Vicario Arjona
Universidad de Salamanca
vicario@usal.es

i. Unas de las flaquezas reconocidas de la noción de contenido semántico surgida de la Nueva Teoría de la Referencia (NTR) tiene que ver con los aspectos cognitivos del significado. Así, ha sido acusada de no dar adecuada cuenta de la dimensión epistemológica del lenguaje. En esta comunicación examinaré una dificultad asociada a una posición surgida de la NTR, la Concepción Híbrida, que trata de atender a esos aspectos cognitivos reconociendo modos de presentación como ingrediente de los pensamientos de los sujetos, pero negando su participación en la semántica (y en la pragmática) de las oraciones.

ii. La noción de contenido, que llamaré *Russellianismo ingenuo* (RI), se caracteriza fundamentalmente por dos tesis:

> *Referencia directa* (RD): El contenido semántico de oraciones simples en las que intervienen nombres propios o deícticos viene dado por una proposición singular. El valor semántico (la contribución a la proposición) de nombres propios y deícticos es meramente el objeto referido.
>
> *Millianismo* (M): El significado de un nombre propio se agota en el referente.

De acuerdo con (RD), las proferencias de (1) y (2) expresan una misma proposición singular, que cabe representar como (1/2p):

> (1) Orwell escribió *Homage to Catalonia*
>
> (2) Eric Blair escribió *Homage to Catalonia*
>
> (1/2p) <escribir- *Homage to Catalonia*, Orwell>

iii. Una cuestión interesante es preguntarnos por la relación entre lenguaje y pensamiento. Naturalmente, al hablar queremos poder decir lo que pensamos. Un *desiderátum* razonable para una teoría semántica sería, pues, que, si uno cree que *p*, pueda aseverar que *p*, y que si uno asevera (reflexiva, sinceramente) que *p*, entonces cree que *p*.

El marco general externista en el que se asienta la NTR provee al partidario de RI de sólidos argumentos para utilizar el aparato de las proposiciones singulares para caracterizar las condiciones de verdad de las creencias. Así, al proferir (1) y

* Este trabajo es parte de proyectos de investigación financiados por el MCeI (HUM2006-09923/FISO; y FFI2008–06164–C02–01) y por la JCyL (VA077A07); y se ha beneficiado de reiteradas y fructíferas discusiones con Manuel García-Carpintero.

creer que Orwell escribió *Homage to Catalonia*, lo que decimos y pensamos tendría las mismas condiciones de verdad. Con ello lograría satisfacerse el desiderátum mencionado.

iv. Es sabido que un hablante puede reaccionar cognitivamente de modo diverso ante oraciones que, como (1) y (2), expresan, según el RI, una misma proposición. Como destacara Frege, un hablante competente puede considerar que (1) es verdadero, y a la vez, suspender el juicio sobre (2) o considerarla falsa. Esta disparidad constituye el *problema del valor* o *significado cognitivo*.

El problema del valor cognitivo apunta a una posible fractura en la relación entre el contenido semántico favorecida por el RI y el contenido del pensamiento. Si un hablante muestra actitudes diversas ante (1) y (2), ello es indicio de que sus pensamientos, sus creencias, de algún modo también son distintos.

Así, sobre la base de este problema puede construirse un famoso argumento en contra del RI: el ***argumento*** *del valor cognitivo*. Consideremos un hablante que juzga que tanto (1) como (2n) son verdaderas, entonces este hablante, digamos Carlos, estaría, según la semántica del RI, aceptando una proposición y su negación, y por tanto tendría creencias contradictorias.

(2n) Eric Blair *no* escribió *Homage to Catalonia*

El crítico de RI puede sostener que si un individuo cree una proposición y su negación, entonces no puede ser racional. Con lo que habría que concluir que Carlos no es racional. Y he ahí el problema, porque intuitivamente no hay razón para considerar que Carlos no sea racional. Imaginemos que Carlos es un vecino de Blair y lector de Orwell, que simplemente no sabe que George Orwell es Blair. Cualquiera de nosotros puede hallarse ante una situación pareja.

v. En principio, la situación provocada por el aparato de las proposiciones singulares no debería sorprender del todo. El mejor modo de entender qué es una proposición singular es considerar que recoge es un *estado de cosas* (posible). No es pues un objeto psicológico, un contenido cognitivo, algo que se capte o sea presente o accesible a la mente; sino algo con relación a lo cual, dos pensamientos de un hablante pueden relacionarse, sin que este sujeto se percate de ello.

vi. El modo clásico (o mayoritario) entre los partidarios del RI de responder al problema del valor cognitivo, y al argumento asociado, es reconocer que hay modos de presentación, modos de conocer proposiciones o sus constituyentes. Así, la relación del sujeto con la proposición está cognitivamente mediada. El modo de responder al argumento contrario al RI pasaría por aceptar que un sujeto puede tener dos creencias, una relacionada con una proposición singular y otra con su negación, y ser, a pesar de ello, racional, siempre que conciba el estado de cosas recogido en las proposiciones de *modos adecuadamente distintos*.

Aunque la maniobra de incorporar modos de presentación es deudora de algunas razones de Frege acerca de la intensionalidad del pensamiento, el partidario de RI mantiene firmes distancias con éste: en particular, los modos de

presentación no son parte del contenido semántico; y tampoco lo determinan. Y ello tanto para creencias como para aseveraciones. En este sentido, tales modos de presentación son semánticamente inertes. A la posición así caracterizada la llamaré, (siguiendo a Richard Heck) *Concepción híbrida* (CH).

vii. No obstante, la simple maniobra de introducir modos de presentación psicológicos es insuficiente a fin de allanar las dificultades que plantea la intensionalidad del pensamiento confrontada con la noción de contenido semántico del RI. En particular, señalaré brevemente que la CH, en sí misma, no permite responder adecuadamente al reto que suponen las oraciones de creencia.

Y es que hablantes (competentes, sinceros, reflexivos,...) pueden considerar que (proferencias de) las oraciones (3) y (4) difieren en valor de verdad, siendo la primera verdadera, y la segunda, falsa. En este sentido, los hablantes se *resisten* a la substitución (en expresión de Braun). Esta intuición de los hablantes es tan firme y generalizada, que constituye un dato muy sólido a favor de que los valores de verdad sean efectivamente como encuentran los hablantes.

(3) Carlos cree que Orwell escribió *Homage to Catalonia*

(4) Carlos cree que E. Blair escribió *Homage to Catalonia*

Naturalmente, eso presenta un serio problema para RI si suplementamos la teoría con un análisis relacional de tales oraciones (por el que "creer" expresa semánticamente una relación diádica entre sujetos y proposiciones, siendo el contenido de la cláusula "que *p*" la proposición expresada por "*p*"). Ya que con (3) y (4) se expresaría la misma proposición singular, y habría de valer en tales contextos el Principio de Substitución (por el que los términos correferenciales son substituibles salva veritate); lo que contradice las intuiciones de los hablantes. Esta dificultad sirve de base para un nuevo argumento (paralelo al mencionado antes) en contra de RI:

(3/4p) <creer, Carlos, <escribir *Homage to Catalonia*, Orwell>>.

Esquema del **argumento** de las **atribuciones de creencia** contra RI:

> (a) Silvia es un agente racional (b) Silvia entiende (3) y (4), y cree que (3) es verdadero y (4) es falso. (c) Si un agente racional entiende (3) y (4), y cree que (3) es verdadero y (4) es falso, entonces cree la proposición expresada por (3) y la negación de la proposición expresada por (4). (d) Luego Silvia cree la proposición expresada por (3) y la negación de la proposición expresada por (4). (e) (3) expresa la misma proposición que (4) [aplicando RI+ *análisis relacional* de las oraciones de creencia]. (f) Silvia cree la proposición expresada por (3) y la negación de esa misma proposición. (g) Ningún agente racional cree una proposición y su negación. (h) Carlos no es un agente racional.

La solución más ortodoxa con la concepción híbrida es una estricta *explicación psicológica* (EP) ofrecida por David Braun [1998, 2002, 2006]. Lo que propone es dar una respuesta unificada al problema de las oraciones de creencia y al del valor cognitivo: "puede extenderse a las intuiciones de los hablantes sobre el valor de

verdad de la oraciones de creencia tales como [(3) y (4)]" la explicación dada por el RI a las intuiciones sobre las oraciones simples (1) y (2) [Braun (1998), p. 572]. Así, lo que explica la resistencia a la substitución de un hablante racional ante las oraciones (3) y (4) es que el hablante concibe de modos adecuadamente diferentes la proposición (3/4p) expresada. En particular, si Silvia es racional cuando cree que (3) es verdadera, y (4), falsa, es porque cree, *de un modo*, la proposición expresada por (3), y cree, *de otro modo, adecuadamente distinto*, la negación de la proposición expresada por (4). Braun rechaza, simplemente, la premisa (g).[1]

viii. La EP dada por Braun al problema de las oraciones de creencia, no funciona adecuadamente. Veámoslo.

a) En este problema la CH adopta una estrategia argumentativa que, en contra de la perspectiva y metodología adoptada por la NTR, no respeta los datos relativos a la práctica lingüística, en este caso las intuiciones de los hablantes, a pesar de ser firmes y generalizadas.

b) La apelación a los modos de presentación como solución a la resistencia a la substitución en las oraciones de creencia suscita serias dudas. Pues mientras que en el caso de Carlos postular una diferencia en sus modos de presentación para la misma persona (Orwell) está plenamente justificado porque Carlos no sabe que Orwell es Blair; en el caso de Silvia no parece estarlo, ya que bien puede suceder que ella *sí* sepa que Orwell es Blair.

La raíz de la insuficiencia explicativa de la EP para dar cuenta del problema de las oraciones de creencia radica en que existe una *asimetría* básica entre los casos con oraciones simples y los casos con oraciones de creencia; y la explicación basada en modos de presentación no logra capturarlo. El contraste entre unos casos y otros es el siguiente: si bien Carlos puede racionalmente creer que las oraciones simples (1) y (2) difieren en el valor de verdad, parece que no puede racionalmente mantener ese juicio si cree que Orwell es Blair. En cambio, por su parte, parece que Silvia sí puede racionalmente mantener su disparidad de juicio sobre los valores de (3) y (4), aun sabiendo que Orwell es Blair. (Recurriendo a la imagen frecuente de concebir los modos de presentación como dossieres, podemos expresar la asimetría de este modo: una vez que un hablante descubre que lo que creía dos objetos es en realidad un sólo objeto, pasa a fusionar o, simplemente *vincular*, los dossieres que tenía para ese objeto. Así, si suponemos que Carlos llega a enterarse de que en realidad Orwell no es otro que Blair, entonces, *salvo irracionalidad*, ese hablante ya no creerá que las oraciones (simples) (1) y (2) difieren en valor de verdad. Sin embargo, un hablante competente, por ejemplo, Silvia, que cree en la identidad de Orwell y Blair, y que tiene vinculados sus modos de presentación o dossieres, puede, aún así, persistir en sus juicios sobre

[1] Mayoritariamente dentro de la CH se explica las intuiciones de los hablantes *pragmáticamente*: los modos de presentación serían parte de un contenido pragmáticamente comunicado. Aunque mostrarlo aquí sobrepasaría los límites de la exposición, la dificultad que afecta a la EP de Braun puede ser extendida a la solución pragmática. Otros partidarios del RI han optado por soluciones semánticas, que reconocen que (3) y (4) difieren en condiciones de verdad.

los valores de las oraciones de creencia (3) y (4) *sin dejar de ser racional* por ello (como de hecho nos pasa a nosotros si juzgamos este caso).[2]

Esta asimetría supone un problema para la EP, pues la explicación que en principio valía para el caso del significado cognoscitivo no parece servir para el de las oraciones de creencia.

Referencias Bibliográficas

Braun, D. (1998), 'Understanding Belief Reports', *The Philosophical Review* 107, pp. 555-95.

— (2002), 'Cognitive Significance, Attitude Ascriptions, and Ways of Believing Propositions', *Philosophical Studies* 108, pp. 65-81.

— (2006), 'Illogical, But Rational', *Noûs* 40, pp. 376-79.

Salmon, N. (1986), *Frege's Puzzle*, Cambridge (Mass.), MIT Press.

Schiffer, S. (1987), 'The 'Fido'-Fido Theory of Belief', *Philosophical Perspectives* 1, pp. 455-80.

Schiffer, S. (2006), 'A Problem for a Direct-Reference Theory of Belief', *Noûs* 40, pp. 361-68.

[2] Schiffer [1987], según creo, fue el primero en destacar este contraste, para oponerse a la posición de Salmon [1986]; aunque posteriormente Schiffer [2006] ha incluido en la crítica a Braun, hace poco por justificar su posición, y no se ocupa de responder a las réplicas de Braun. En "No tantos modos de creer proposiciones" (próx.) ofrezco una justificación adecuada, y respondo a la defensa a la objeción elaborada por Braun.

La intención y la representación espacial del movimiento*

Marta Vidal Perera
Universitat Autònoma de Barcelona
Marta.Vidal.Perera@uab.cat

La diferencia entre un simple movimiento y una acción corporal suele explicarse a partir del hecho que sólo en la última encontramos una intención, aquel estado mental por el que un agente hace el movimiento corporal. Que en toda acción haya una intención no quiere decir que toda intención sea parte constitutiva de una acción[1]. Por ejemplo, la intención *el año que viene comeré más sano* no es algo por lo que un agente pueda estar haciendo un movimiento corporal, ni parece poder serlo. Para diferenciar las intenciones constitutivas de una acción, es decir, aquellas por las que un agente realiza un movimiento, de las intenciones entendidas en el sentido cotidiano, algunos autores dan a éstas un nombre especial: 'intenciones-en-la-acción' (Searle) o 'intenciones-M' (Pacherie).

Mi objetivo es abordar un tema clave en la caracterización de estas primeras intenciones constitutivas de una acción (que aquí llamo simplemente 'intenciones'): la representación espacial que hacen del movimiento. El estudio de este aspecto de la intención forma parte del proyecto más amplio de conceptualizar la relación entre los estados mentales del agente y sus movimientos corporales.

La experiencia de la acción

La intención constitutiva de la acción es clave para explicar la experiencia de la acción. Es comúnmente aceptada la tesis según la cual el contenido de la experiencia de la acción es la intención [cf. Searle (1983), Pacherie (2008), Peacocke (2003)]. Esta tesis da cuenta de lo siguiente (y éste es un dato en su favor): dos acciones son experimentadas distintamente si, aún compartiendo el mismo movimiento, responden a intenciones distintas. La acción de coger el vaso y la de comprobar la movilidad del brazo derecho pueden compartir idéntico movimiento y, sin embargo, el contenido de su experiencia es distinto.

Nombro como 'EA(*int*)' la tesis por la cual el contenido de la experiencia de la acción es el contenido de la intención: 'EA' por experiencia de la acción, e '*int*', entre paréntesis, como contenido de la experiencia, que aquí es la intención.

* Este trabajo ha sido posible gracias a la beca FPU otorgada por el Ministerio de Educación y se ha beneficiado de la financiación del Ministerio de Ciencia e Innovación a través del proyecto de investigación FFI2008-06164-C02-02 y de la ajuda de la Generalitat de Catalunya al Grup d'Investigació en Epistemologia i Ciències Cognitives (GRECC) SGR2009-1528.

[1] En adelante utilizaré el término 'acción' para referirme a la acción corporal.

El movimiento en la experiencia de la acción

¿Como explica EA(*int*) la experiencia del movimiento que se realiza en la acción (p. ej., la experiencia de estar dirigiendo el brazo hacia la mesa para coger el vaso)? En un principio la propiocepción es el sentido que tiene la función de informar sobre la posición de los músculos y los movimientos realizados sin necesidad de contacto visual o táctil con el cuerpo.

EA(*int*) defiende, sin embargo, que el movimiento que se cree haber experimentado no es el que procede de la propiocepción, sino el representado en la intención, dando por supuesto que las intenciones siempre representan un movimiento. De hecho, el contenido de una intención representa qué tiene que suceder para considerarla satisfecha (sus condiciones de satisfacción) y parte de lo que tiene que suceder es que el agente haga algo con su cuerpo. Si no fuese así, el sujeto podría considerar satisfecha una intención sin que fuese condición de satisfacción algún movimiento corporal, y esto no parece una opción plausible de considerar satisfecha una intención.

Los experimentos de Marc Jeannerod (2003) y Benjamin Libet (1985) suelen utilizarse para argumentar que la información sobre el movimiento procede principalmente de la intención y no de los datos propioceptivos (cf. Pacherie 2008). En el experimento de Jeannerod los sujetos no experimentaban los movimientos correctores inintencionados que realizaban en la acción. Si la experiencia del movimiento procediese de la propiocepción probablemente los sujetos habrían detectado ese movimiento corrector. Por su lado, los resultados de Libet muestran que la realización del movimiento en una acción parece experimentarse 200 milisegundos antes de que el músculo inicie el movimiento. El movimiento que el sujeto cree experimentar no puede proceder, por tanto, de la propiocepción. Libet propone, en cambio, que procede del RP, unos procesos neuronales eléctricos previos a la acción, que identifica con la intención. Así, la intención y no el movimiento explicaría la presunción de experiencia del movimiento cuando actuamos.[2]

La Condición de Coherencia y el experimento de Marcel

Si para EA(*int*) la fuente de la presunción de la experiencia del movimiento es la intención, una condición para su plausibilidad ha de ser que la intención sea coherente con el movimiento realizado. De algún modo la experiencia de la acción debe servir como indicador del movimiento y esto sólo puede suceder si se da esta coherencia. Hay coherencia cuando los movimientos que representa la intención es alguno de los movimientos realizados (o todos ellos). Si la intención representase

[2] Aún así, la experiencia procedente de la propiocepción parece desarrollar algún papel. En el experimento de Jeannerod cuando la desviación de la mano era superior a 14º, los sujetos percibían los movimientos correctores. Pero, para EA(*int*), aunque la propiocepción indique posibles desvíos respecto al movimiento representado en la intención, la experiencia de qué movimiento realizamos procede en un primer momento de la intención.

un movimiento del brazo a la derecha y el movimiento realizado fuese sólo el de levantar el brazo, no se cumpliría la condición. Peacocke (2003) defiende algo similar a esa condición de coherencia –que llamo 'CC'– al afirmar que los intentos deben implicar las especificaciones motoras de la acción.

Presento la coherencia entre intenciones y movimiento como condición de plausibilidad de EA(*int*), en tanto que EA(*int*) es compartido por la mayoría de autores que tratan la experiencia de la acción. Sin embargo para hipótesis más fuertes, que defienden que las intenciones no sólo son el contenido de la experiencia de la acción, sino que causan, además, el movimiento (paradigmáticamente la propuesta de Searle), CC parece aún más necesaria: sería difícil concebir que la intención de mover el brazo a la derecha causase el movimiento del brazo hacia arriba.

Anthony Marcel (2003) presenta, sin embargo, una serie de experimentos que ponen a CC en una situación grave. Los sujetos, tras una estimulación vibrotáctil del brazo derecho, creían tener el brazo perpendicular al torso, aunque lo tenían paralelo al mismo (el brazo estaba cubierto y no podían verlo). Para dirigir el brazo a un punto que se les indicaba debían moverlo separándolo del torso hacia el exterior, de izquierda a derecha [↗], aunque, según la posición ilusoria del brazo, el movimiento tendría que ser el opuesto, hacia el torso, de derecha a izquierda [↙]. La intención que los sujetos decían tener era la de mover el brazo al punto indicado mediante el movimiento incorrecto. Sorprendentemente, aun teniendo esta intención, realizaban el movimiento correcto. Parece, entonces, que deba aceptarse la posibilidad de incoherencia entre el movimiento representado por la intención y el realizado.

Una manera de mantener la coherencia entre las intenciones y el movimiento (y, por tanto, de continuar descansando EA(*int*) en CC) la proporciona Peacocke (2003), que, al comentar el experimento de Marcel, propone que la experiencia de la acción presenta dos contenidos. Uno, al que llama 'α', es del tipo *estoy moviendo el brazo de derecha a izquierda hasta la luz*. El otro, al que llama 'β', es del tipo *estoy moviendo el brazo hasta ese punto*. Este contenido β permite mantener la coherencia entre la intención y el movimiento puesto que no especifica un movimiento del brazo a la derecha o a la izquierda.

Así, el esquema de la experiencia de la acción sería éste:

intención	contenido α	(*mover el brazo de derecha a izquierda hasta la luz*)	
	contenido β	(*mover el brazo hasta ese lugar*)	> CC
datos propioceptivos		(*movimiento del brazo de izquierda a derecha*)	
EA	contenido α	(*mover el brazo de derecha a izquierda hasta la luz*)	

Peacocke no da cuenta sin embargo de por qué para el sujeto el contenido α no es equivalente al contenido β, ni explica por qué los sujetos creen experimentar una intención incorrecta. Mi aportación al debate radica en desarrollar una posible explicación a estos hechos: el contenido α representa espacialmente el movimiento de modo alocéntrico, mientras que el contenido β lo representa de modo egocéntrico.

Representaciones egocéntricas y representaciones alocéntricas

Una representación egocéntrica es aquella que requiere necesariamente términos relativos al sujeto para representar el entorno y sus objetos (p. ej., enfrente, a mi derecha o allí). Una representación alocéntrica, en cambio, los localiza objetivamente, sin relación a un sujeto, como un punto en un mapa.

Para dar cuenta de la subjetividad propia de las representaciones egocéntricas su contenido se enuncia en términos de disposiciones a la acción: tener una representación egocéntrica de un objeto o un lugar consiste en estar dispuesto a hacer determinadas acciones (cf. Evans 1982, Peacocke 1983, Campbell 1994). La explicación es ésta: sólo se consigue dar cuenta de una relación básicamente subjetiva con el entorno si el origen de la representación es una imagen de la que el sujeto tiene conocimiento directo y no observacional; y una imagen así sólo puede relacionar al sujeto con el entorno a partir de su capacidad para actuar en ese entorno.

El contenido *mover el brazo a ese punto*, puesto que presenta una localización egocéntrica (*ese punto*), representa posibles acciones del brazo. Las condiciones de satisfacción de ese contenido son, pues, la actualización de una de las acciones posibles. Estas condiciones de satisfacción son coherentes tanto con el movimiento del brazo a la derecha como con el movimiento a la izquierda. Esto no quiere decir que cualquier acción satisfaga esta representación. Sólo se satisfará si el sujeto realiza alguna de las acciones posibles.

Evans señala un rasgo de las representaciones egocéntricas: sólo pueden formar parte de un sistema de razonamiento y de pensamiento abierto a la conceptualidad si tienen conexión con las coordenadas del espacio objetivo, es decir si representan alocéntricamente el entorno. La manera como una representación egocéntrica tal como *dirigir el brazo a ese punto* se sitúa en un espacio objetivo es mediante un razonamiento como éste:

> Si ese punto [localizado egocéntricamente] está en π [un punto del espacio alocéntrico] y mi brazo [localizado egocéntricamente] está en ρ [un punto del espacio alocéntrico], tengo la intención de mover el brazo de derecha a izquierda.

El consecuente de este condicional es lo que Peacocke llama 'contenido α'. Este contenido α depende de una identificación en el antecedente, que puede basarse en una información errónea, como sucede en el experimento de Marcel por la ilusión creada por la estimulación vibrotáctil. Esto hace que aun siendo correcta la

representación egocéntrica, el input del sistema de pensamiento y razonamiento consciente pueda ser erróneo.

Conclusión

He presentado aquí una manera que en la que EA(*int*) podría mantener CC. Consiste en considerar que el contenido de la intención representa en un primer momento egocéntricamente el espacio y, mediante un razonamiento que permite que forme parte de un sistema de pensamiento consciente, lo representa alocéntricamente. En el desarrollo de una acción, la representación egocéntrica mantiene la coherencia con el movimiento. En condiciones normales, la representación alocéntrica también la mantiene. Sin embargo al basarse en un razonamiento con información sobre dónde se sitúan el sujeto y sus miembros, esta representación no mantiene infaliblemente la coherencia con el movimiento.

Referencias bibliográficas

Campbell, J. (1994), *Past, space, and self*, Cambridge, MIT.
Evans, G. (1982), *The Varieties of Reference*, Oxford, Clarendon Press.
Jeannerod, M. (2003), 'Consciousness of Action and Self-Consciousness: A Cognitive Neuroscience Approach', en Roessler y Eilan (eds.) (2003).
Libet, B. (1985), 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioral and Brain Sciences* 8, pp. 529–66.
Marcel, A. (2003), 'The Sense of Agency: Awareness and Ownership of Action', en Roessler y Eilan (eds.) (2003).
Pacherie, E. (2008), 'The phenomenology of action: A conceptual framework', *Cognition* 107, pp. 179-217.
Peacocke, Ch. (1983), *Sense and Content*, Oxford, Clarendon Press.
— (2003), 'Action: Awareness, Ownership, and Knowledge', en Roessler y Eilan (eds.) (2003).
Roessler, J. y N. Eilan (eds.) (2003), *Agency and Self-Awareness*, Oxford, Oxford Univeristy Press.
Searle, J. (1983), *Intentionality*, Cambridge, Cambridge University Press.

Algunas ideas para "re-actualizar" los argumentos escépticos

Javier Vilanova Arias
Universidad Complutense de Madrid
vilanova@filos.ucm.es

Introducción

En la antigüedad clásica el debate realismo-escepticismo (o escepticismo-dogmatismo) es una cuestión de grado: los argumentos escépticos van dirigidos a probar que el error es tan o más probable que el acierto (de dónde se seguiría que nuestras pretensiones de conocimiento no son legítimas), y para ello el escéptico recurre a los mismos tipos de evidencias y creencias justificadas que el realista utiliza. De ahí los diez "modos" de Sexto Empírico, en los que de cada argumento se concluye una mayor probabilidad de error a partir de la presencia de algún tipo específico de contradicción entre evidencias. En la modernidad el debate realismo-escepticismo es un cuestión de todo-nada: los argumentos escépticos van dirigidos a probar que es posible (lógicamente posible) el error, y para ello el escéptico no puede utilizar creencias que se justifican en base al mismo tipo de evidencias que el realista utiliza. Esto en un sentido concede una ventaja al escéptico (que ha de probar un enunciado más débil), pero en otro sentido supone una desventaja (los recursos argumentales se ven seriamente restringidos). A la larga, las reglas de juego de la modernidad han degenerado el debate hasta convertirlo en una suerte de diálogo de sordos: el realista recusando las evidencias del escéptico por espurias o auto-contradictorias, el escéptico denunciando que las justificaciones para la creencia que presenta el realista son compatibles con el error. De ahí la actitud prevalente en la contemporaneidad respecto al debate escepticismo-realismo: éste es un debate absurdo, un pseudo-problema creado por los filósofos profesionales y totalmente desconectado con preocupaciones epistemológicas reales del científico o del hombre de la calle (el Wittgenstein de "Sobre la Certeza" es un ejemplo paradigmático de esta actitud).

En este trabajo pretendemos reavivar el debate escepticismo-realismo, desde la convicción de que, si se encara desde un marco más próximo (aunque no idéntico) al de la antigüedad clásica, se puede recuperar el contacto con intereses y problemas reales. Para ello, nos apoyaremos en ciertas consideraciones que hace George E. Moore en 1941 en torno a los argumentos escépticos que pretenden probar la posibilidad del error, si bien las consecuencias que extraeremos de su análisis serán muy distintas (Moore presenta su análisis como un contra-argumento, nosotros lo tomaremos como una revisión del argumento escéptico).

El argumento escéptico modal de Moore

El argumento escéptico que Moore discute es el siguiente:

El argumento escéptico modal (o argumento de Carneades)

(P.1.) Sean cuales sean mis evidencias para creer p, esas evidencias son compatibles con no p.

(P.2.) Si mis evidencias para creer p son compatibles con no p, entonces no sé que p.

Por lo tanto, (p.3.) No sé que p.

La expresión "mis evidencias son compatibles con no p" puede parafrasearse como: "es posible que se den las evidencias de que dispongo para justificar mi creencia de que p y que la proposición sea falsa". Así, el argumento modal se puede parafrasear como:

(P.1.) Sean cuales sean mis evidencias para creer p, es posible que se den esas evidencias y p sea falso.

(P.2) Si es posible que mis evidencias para creer p sean verdaderas y p sea falso, entonces no sé que p.

Por lo tanto, (P.3) No sé que p.

El contra-argumento de Moore consiste en acusar al escéptico de estar cometiendo una falacia de equivocidad, aprovechándose de la ambigüedad de las nociones modales manejadas por el escéptico. En realidad, estaríamos ante una triple ambigüedad en la forma de entender la expresión "es posible":

(1) -in sensu diviso/in sensu composito (o de re/de dicto):

Posible(P y ¬E)

(Posible P) y ¬E

Aplicada al caso epistémico, siendo J una justificación para creer que p, el escéptico usaría en la premisa C1 "es posible" como "J y es posible que no p", mientras que en la premisa C2 entendería "es posible" como "es posible J y no p a la vez". Hacer esta distinción es importante, ya que, por poner un ejemplo, podemos conceder al escéptico que una proposición p que versa sobre el mundo externo es contingente y por lo tanto podría haber sido falsa, pero aun no le habremos concedido que es posible al mismo tiempo que la proposición sea sabida y que p sea falsa.

(2) -para todos a la vez/para cada uno:

Para todo e (Posible ¬e)

Posible (para todo e (¬e)

En el caso que nos ocupa, siendo J1, J2…Jn mis justificaciones para creer que p, en la premisa P1 el escéptico estaría usando "es posible" en el sentido de "es posible, para cada una de las justificaciones J, que J sea verdadero y p falso al mismo tiempo", y en la premisa P2, como "es posible que la conjunción de todas las evidencias J sea verdadera al mismo tiempo que p es falso". Una vez detectada esta ambigüedad pierde bastante de su fuerza el recurso a errores de percepción

pasados (espejismos, sueños, efectos ópticos, etc...) que hace el escéptico, sobre todo cuando las proposiciones presuntamente sabidas son, como las que elige Moore ("tengo una mano", "existe un mundo externo"), apoyadas por una gran cantidad de evidencias recogidas a lo largo de un extenso periodo de tiempo como lo es la propia vida. Pues si bien es posible imaginar para cada una de las percepciones que sirven de evidencias que podría tratarse de un error, es más difícil idear cómo podrían haberlo sido todas ellas.

(3) -posible naturalmente (o naturalmente probable)/lógicamente:

Esta es, a mi juicio, la distinción más fundamental. Ésta es la que hay entre una noción de posibilidad absoluta, o "posibilidad lógica", y otras nociones de posibilidad a las que de momento identificaré como una sola, y denominaré "posibilidad natural" o "posibilidad física". Llevando la distinción al argumento de Carneades, resultará que podemos conceder tranquilamente al escéptico la premisa C1 si por posible entendemos "lógicamente posible", sin que ello nos haga desdecirnos de nuestras pretensiones de conocimiento. En efecto, es lógicamente posible que yo haya seguido tales y tales reglas epistémicas, haya obtenido tales y tales evidencias para creer p, y sin embargo p sea falso. Esto es algo que el propio Moore concede al escéptico, como hemos dicho, pero este hecho podría no ser suficiente para poner en cuestión nuestras reglas epistémicas. En efecto, para Moore y mucha otra gente (yo entre ellos) la cuestión no es si es posible **lógicamente** el error. El error es siempre lógicamente posible, ya que lo único que es lógicamente imposible son cosas como "p y no p" o "p y q implica no p". Por ello, decir que es lógicamente posible el error es tan trivial como decir que "p y no p" es imposible, o que "p o no p" es necesario, y completamente irrelevante cuando discutimos un problema epistemológico (como es irrelevante, por ejemplo, afirmar que es lógicamente posible que un móvil supere la velocidad de la luz cuando discutimos problemas de física). Lo que hay que dirimir, la cuestión importante en epistemología, es si nuestras maneras de relacionarnos causal y perceptivamente con nuestro entorno junto con las maneras en que las cosas ocurren en nuestro entorno (las leyes físicas) y las maneras en que discurren nuestros propios procesos mentales abren o no un ámbito de posibilidades en el que sean cuales sean mis reglas epistémicas, mis datos sensoriales y mi forma de procesarlos hay siempre una situación en que se da el error (en que p es falso). Por supuesto, el escéptico puede no compartir este punto de vista, y con él otros filósofos, pero entonces deberá darnos razones adicionales que nos persuadan para abandonar nuestro punto de vista y adoptar el suyo.

Revisemos el argumento de Carneades a la luz de las ambigüedades detectadas. En las versiones más pueriles, el escéptico estaría cometiendo una falacia de equivocidad, usando "posible" en sentidos distintos:

> En premisa p1: como "es para cada una posible lógicamente in sensu diviso".
>
> En premisa p2: como "es para todos naturalmente posible in sensu composito".

En versiones más sofisticadas la ambigüedad sólo afectaría a una o dos de las tres dicotomías señaladas. En cada caso, el error del escéptico sería el mismo: falla en probar la premisa p1 leyéndola según la misma noción de posibilidad que requerimos en la lectura de la premisa 2.

El patrón argumental modal escéptico

Esto en lo que respecta al argumento modal de Moore (que yo he reconstruido más o menos libremente a partir de mi propia exégesis de los textos de Moore), entendido como un contra-argumento del realista versus el argumento modal del escéptico. En lo que sigue intento re-elaborar el argumento para poder aplicarlo a situaciones reales en las que alguien tiene una duda razonable sobre sus presuntos conocimientos. Ahora bien, para hacer el esquema argumental efectivo en tales contextos debemos dotarle de mayor flexibilidad. Ya hemos visto que combinando los miembros de las dicotomías asociadas a las tres ambigüedades de la palabra "posible" obteníamos versiones diferentes del argumento. Añadiendo nuevos miembros a esas ambigüedades obtendremos una flexibilidad todavía mayor. Para empezar, introduciendo nuevos tipos de posibilidad además de la posibilidad "física" y la posibilidad "lógica" que sean relevantes en algunos contextos. Podemos o no introducir consideraciónes fisiológicas, históricas, tecnológicas, geográficas, contextuales, etc… Además, no es necesario que "todos" las justificaciones sean espurias (ni tampoco, como algún filósofo ha proclamado exageradamente, basta con que lo sea uno). Dependiendo del contexto y del *standard* de certeza buscado (no es el mismo, por ejemplo, en una discusión entre amigos que en un debate científico) la proporción de errores admisible puede variar. Por último, si introducimos escalas (no necesariamente numéricas) de posibilidad o probabilidad, el esquema será todavía más flexible y por lo tanto más aplicable. Llegamos, de este modo, a la familia de nociones de posibilidad que surge de saturar las tres variables mencionadas con valores tomados de conjuntos muy amplios:

Es

(muy, bastante, algo, poco, casi nada…)

posible

(físicamente, , biológicamente, antropológicamente, históricamente, en el contexto actual, etc…)

para

(muchos, bastantes, suficientes, unos cuantos, casi todos…)

Podemos así, hablar de un patrón argumental (llamémosle "patrón argumental escéptico modal") que produciría una gran familia de argumentos concretos surgidos al interpretar la expresión "es posible" en las premisas del argumento escéptico modal según una de las lecturas descritas previamente. No todos ellos, como defenderé, son falaces, y algunos pueden ser cogentes en situaciones de duda reales. Las realizaciones del esquema argumental que pretenden producir dudas globales (dudas sobre toda una fuente de evidencias o un sistema epistémico

completo) en situaciones no anómalas a partir de una débil posibilidad de error (sobre, por ejemplo, todo el conocimiento perceptivo de un individuo normal) son en general poco o nada cogentes. Estos realizaciones son, en general, los argumentos escépticos que discuten los modernos (Descartes, Hume y compañía) y que los contemporáneos tildan de nada interesantes o simples sinsentidos (Carnap, Wittgenstein y otros muchos). Los argumentos escépticos que discuten los clásicos (Pirrón, Sexto Empírico, Carneades y sus antagonistas) intentan producir dudas parciales en situaciones normales que siendo acumuladas lleguen a producir una duda global. En mi opinión, los argumentos escépticos clásicos, si bien mucho más interesantes que los que discuten los modernos, tampoco son cogentes. Sin embargo, esto no se cumple para las realizaciones del esquema argumental que producen dudas parciales en situaciones normales, por ejemplo en torno a las creencias asociadas a un tipo de acciones epistémicas (por ejemplo, ver el aspecto del cielo para saber si va a llover), o a una fuente de evidencias más específica (por ejemplo, una página de internet). Ni tampoco se cumple para algunas dudas globales que tienen lugar en situaciones anómalas, como por ejemplo en contextos de crisis, ya sean científicas, culturales o de valores, o para individuos con desordenes perceptivos graves (según su propio testimonio el matemático y premio nobel John Nash llegó a ser consciente de sus delirios psicóticos siguiendo argumentos que se parecen mucho al de Carneades y al de Protágoras), o en regímenes políticos super-manipuladores de la información, etc... En todos estos casos, una de las cosas que el individuo hace es dilucidar hasta qué punto es probable que una parte importante de sus creencias del tipo X sean falsas. Existen, además, una serie de falacias típicas que alguien puede cometer al seguir el argumento de Carneades. No entraré en detalles ahora, pero la mayor parte de ellas tienen que ver, o bien con una elección de los valores para las variables del vector "posible" en la premisa P2 que no se ajusta al tipo de certeza adecuado al contexto (por ejemplo, exigir un alto grado de posibilidad en un contexto de discusión familiar), o bien con proporcionar justificaciones para la premisa P1 desambigüada con determinados valores para las variables del vector "posible" que no se corresponden con los valores con los que se ha desambigüado las premisa C2.

Meteorological sentences, unarticulated constituents and relativism

Dan Zeman
Universitat de Barcelona, LOGOS
dan_zeman@yahoo.com

One of the debates that has taken center stage in contemporary philosophy of language is that between *literalism* – roughly, the thesis that the literal, linguistic meaning of a sentence is enough to determine the truth conditions of that sentence as used in a certain context, and *contextualism* – the thesis that besides the linguistic meaning of a sentence there are other – contextual, pragmatic – factors that contribute to the truth-conditions of that sentence as used in a certain context. The contribution of those pragmatic factors consists in providing unarticulated constituents that enter in the truth-conditions of the sentence.

In my paper I focus on the particular case of meteorological sentences such as "It is raining", as they are highly representative for the debate between contextualism and literalism. Specifically, in what follows I will have a look on the exchange between Jason Stanley and Francois Recanati pertaining to the debate mentioned. The exchange revolves around the issue whether in a sentence like "It is raining" the location of rain is part of the logical form of the sentence or is an unarticulated constituent provided by the context. As widely known, Recanati defends an unarticulated constituent analysis, whereas for Stanley "any contextual effect on truth-conditions that is not traceable to an indexical, pronoun, or demonstrative (…) must be traceable to a structural position occupied by a variable" (Stanley, 2000: 401). Despite their disagreement in the "It is raining" case, something that the two authors agree upon is the formulation of an unarticulated constituent. Here is what it means to be an unarticulated constituent, according to Stanley:

> *x* is an unarticulated constituent of an utterance *u* iff (1) *x* is an element supplied by context to the truth-conditions of *u*, and (2) *x* is not the semantic value of any constituent of the logical form of the sentence uttered. (Stanley, 2000: 410)

However, even if armed with this definition, it is not clear when a given *x* is an unarticulated constituent or not, for clause (2) of the definition is where the debate actually begins. Therefore, both authors have been keen to provide alleged criteria for unarticulatedness. Thus, Recanati has provided what he has called

> ***The Optionality Criterion***: Whenever a contextual ingredient of content is provided through a pragmatic process of the optional variety, we can imagine another possible context of utterance in which no such ingredient is provided yet the utterance expresses a complete proposition. (Recanati, 2004),

whereas Stanley has proposed

> ***The Binding Criterion***: A contextually provided constituent in the interpretation of a sentence S is articulated whenever the argument role it fills can be intuitively "bound", that is, whenever what fits that role can be made to vary with the values introduced by some operator prefixed to S. (Stanley, 2005)

It is easy to show that both criteria have troubles; specifically, both criteria are liable to overgeneralize, thus yielding wrong results. Recanati's illustration of the Optionality Criterion is the well-know weatherman case; but it seems easy to come up with similar examples for which such contexts could be constructed, yet we are not willing to treat the expression as having one argument less in its logical form (for example, the predicate "kiss"). As for Stanley's criterion, a host of examples have been offered showing that, were the criterion right, we would end up postulating variables in a much bigger number of cases than expected. (One example to this effect is the sentence "Everywhere Janie went, she danced", due to Sennett (2008)).

However, something of more importance comes up when we look at the main challenge that Stanley has raised for the contextualist: namely, to account for cases in which binding by a second-order quantifier occurs. Famously, sentences like

> Every time John lights a cigarette, it is raining,

spawn trouble for the contextualist, since the resources of the contextualist theory are too poor to yield the required reading. Given that the semantic clause for "rain" is something along the following lines:

> Den("rains") relative to a context c = that function f that takes $<t, l>$ to True if it is raining at t in l, where l is the contextually salient location in c, takes $<t, l>$ to False if it is not raining at t in l, where l is the contextually salient location in c, and it undefined otherwise,

the contextualist analysis of the sentence "It is raining" is

> "It is raining (t)" is true in a context c iff the denotation of "rains" takes $<t, l>$ to the True, where l is the contextually salient location in c.

But this analysis doesn't have enough resources to handle a more complicate sentence like "Every time John lights a cigarette, it is raining". One reading of this sentence (and the relevant one here) is that for every time t at which John lights a cigarette, it rains at t at the location in which John lights a cigarette at t. But, the only available rendering of the sentence for the contextualist is

> For every time t at which John lights a cigarette, the denotation of "rains" takes $<t, l>$ to the True, where l is the contextually salient location in the context of utterance of "Every time John lights a cigarette, it is raining",

which is not the required reading. Since an analysis according to which the location of rain is part of the logical form of such sentences has no problem with such examples, it should be preferred over an unarticulated constituent analysis.

This conclusion has far-reaching consequences. For once Stanley has established that the location of rain must be part of the logical form of a sentence like "Every time John lights a cigarette, it is raining", it also must appear in the unembedded sentence ("It is raining"). The argument, known as *the binding argument*, can be put as follows:

1. Unarticulated constituent theorists say that in the simple statement "It is raining", the location of rain is unarticulated.
2. In "Every time John lights a cigarette, it is raining", binding occurs: the location of rain varies with the values introduced by the quantifier "every time John lights a cigarette".
3. There is no binding without a bindable variable.
4. Therefore, "It is raining" involves a variable for the location of rain.
5. It follows that the unarticulated constituent theorist is mistaken: in the simple statement "It is raining", the location of rain is articulated. It is the (contextually assigned) value of a free variable in logical form, which variable can also be bound (as in the complex sentence "Every time John lights a cigarette, it is raining").

This anti-contextualist argument, however, can be escaped. Recanati has answered the challenge posed by Stanley by employing *variadic functions* (functions from properties to properties having the role of decreasing or increasing the adicity of a predicate), which allow him to avoid the conclusion of the binding argument. The formal machinery involves defining a general increasing variadic operator, **Circ**, and a host of specific operators of the same kind, for specific circumstances (like location, time, etc.). Let us see how this works with a simple example. In a sentence like "John eats in Paris", the phrase "in Paris" is treated as a variadic locational operator operating in the predicate "eat", transforming it from a one-place predicate into a two-place predicate; formally,

$$\mathbf{Circ}_{\text{location: Paris}}\,(\text{Eats (John)}) = \text{Eats\_in (John, Paris)}.$$

The effect of the variadic operator in twofold: on one hand, it increases the adicity of the predicate applied to ("eat"); on the other, it provides the value for the new created argument place (the value in this case being Paris). Applied to the troublesome example "Every time John lights a cigarette, it is raining", the idea is to treat the quantifier ("every time John light a cigarette") as a variadic operators having a double role: that of creating an extra argument place in the predicate it applies (in this case, "rain") and of providing a range of values for the new created argument. So, by employing variadic operators, Recanati can resist the conclusion of the binding argument and the unarticulated constituent analysis of "It is raining" is saved. Thus, my conclusion is that, even if his criterion for unarticulateness is flawed, Recanati's rejection of the conclusion of the binding argument gives him a dialectical advantage over Stanley.

The claim I want to make in the this second part of the paper is that the employment of variadic functions could be useful for *relativism* about a series of discourses, such as predicates of personal taste, knowledge attributions or

epistemic modals. Relativism, as I use the term in the paper, is the view that the truth-value of a sentence varies with the circumstances of evaluation against which the sentence is to be evaluated in such a way that the sentence could be true as uttered in one context and false in another – without any change in the proposition expressed by the utterances of the sentence in the two contexts. This view goes hand in hand with the idea that context has not only a *content-determinative* role, but also a *circumstance-determinative* role (in the terms of MacFarlane (2009)); in other words, context's role in not only to provide elements that end up in the content of sentences uttered, but also to provide the circumstances against which such sentences have to be evaluated. In some cases, as that of meteorological sentences (as seen above), it is clear that context contributes a circumstance rather than an element in the content. This phenomenon has been already brought to the fore in John Perry's (1986) Z-landers story. My claim is that something similar happens in the case of sentences containing predicates of personal taste, epistemic modals and epistemic terms like "know".

Here is an illustration of the advantage that the relativist can gain by employing variadic functions is his semantic machinery. Take, for example, a sentence like "Avocado is tasty". Although no one has applied the binding argument to predicates like "tasty", it doesn't seem hard to come up with bound readings involving "tasty" – sentences in which the person for which something is said to be tasty varies with a quantifier. The following example might be used against the relativist:

> The zoo keeper brought the food. Every animal got something tasty.

whose logical form is

> Every animal from the zoo *x* got some food *y* such that *y* was tasty for *x*.

The relativist has a ready answer to any attempt to apply to binding argument in this case. Following Recanati, the relativist can treat the quantifier in the above examples as a variadic operator that both increases the adicity of the predicate "tasty" and provides a range of values for the new created argument. What the relativist has to do is to define a specific variadic operator $\mathbf{Circ}_{\text{taste}}$ which for a simple sentence like "Avocado is tasty for John" functions formally in the following way:

> $\mathbf{Circ}_{\text{taste: John}}$ (tasty(avocado)) = tasty_for (avocado, John).

This result clearly supports relativism. Given a (I take it) uncontroversial principle, formulated by Recanati as

> ***Distribution***: The determinants of truth-value distribute over the two basic components truth-evaluation involves: content and circumstance. That is, a determinant of truth-value, e.g. a time, is *either* given as an ingredient of content *or* as an aspect of the circumstance of evaluation. (Recanati, 2007),

the relativist is free to treat simple sentences as "Avocado is tasty" as lacking an argument for a subject in their logical form, and take context as providing the circumstance rather than an element in the content of such sentences – which, in turn, allows him to safely appeal to the disagreement data that initially motivated

the view. Similar results are available in the case of epistemic modals and knowledge attributions.

References

MacFarlane, J. (2009), 'Nonindexical Contextualism', *Synthese* 166, pp. 231-50.

Perry, J. (1986), 'Thought without Representation', *Supplementary Proceedings of the Aristotelian Society* 60, pp. 137-52.

Recanati, F. (2004), *Literal Meaning*, Cambridge:, Cambridge University Press.

— (2007), *Perspectival Thought: A Plea for Moderate Relativism*, Oxford, Oxford University Press.

Sennet, A. (2008), 'The Binding Argument and Pragmatic Enrichment, or, Why Philosophers Care Even More than Weathermen about 'Raining'', *Philosophy Compass* 3/1, pp. 135-57.

Stanley, J. (2000), 'Context and Logical Form', *Linguistics and Philosophy* 23, pp. 391-434.

— (2005), 'Review of François Recanati, *Literal Meaning*', *Notre Dame Philosophical Reviews*, retrievable on-line at <http://ndpr.nd.edu/review.cfm?id=3841>.

Sección C

Filosofía y metodología de la ciencia

Clases naturales

Sebastián Álvarez Toledo
Universidad de Salamanca
sat@usal.es

> "... we shall at least be freed from the vain search for the undiscovered and undiscoverable essence of the term species". (Darwin, *The Origin of Species*)

No es lo mismo situar a los osos polares en el grupo de los mamíferos que agruparlos con los animales blancos. Los mamíferos constituyen un grupo con una justificación y una solidez de las que carece el gratuito conjunto de los animales blancos. Esta diferencia es la que expresa la división entre clases naturales y no naturales. Sin embargo, la noción de clase natural dista mucho de resultar clara y las discusiones acerca de su significado, e incluso de la existencia de tales clases, vienen suscitando desde hace décadas un notable interés en metafísica, filosofía de la ciencia y filosofía del lenguaje.

Desde un punto de vista fuertemente realista, clasificar de modo natural un conjunto de organismos, minerales o estrellas consiste en reproducir la estructura en que realmente están organizados en la naturaleza, trinchar la naturaleza por sus propias articulaciones. Hay quienes, desde esta concepción, defienden que la pertenencia de un individuo a una clase natural viene determinada por una propiedad *esencial* que, al mismo tiempo, lo define. Esta propiedad se convierte en condición necesaria de pertenencia a la clase. A veces se trata de una propiedad esencial microestructural que la ciencia nos da a conocer (como el número atómico de un elemento o la composición química de un compuesto), y que puede llegar a constituir una condición necesaria y suficiente. Según Ellis (2001, 145-150), las propiedades esenciales de las clases naturales son el soporte en que se sustentan las leyes de la naturaleza y su necesidad metafísica.

Sin embargo, este esencialismo debe hacer frente a serias dificultades. Por ejemplo, la composición química de una sustancia resulta a veces insuficiente para clasificarla, como muestra el caso de los isómeros, sustancias compuestas por los mismos elementos y en la misma proporción pero que exhiben distintas propiedades; o el hecho de que una simple molécula de agua no exhiba las propiedades termodinámicas características del agua, que son fruto de interacciones moleculares. Y en biología, aunque se acepta generalmente que las especies son ejemplos típicos de clases naturales, existen actualmente diversos criterios para identificarlas. Junto a la tradicional caracterización de las especies por criterios morfológicos, coexisten la concepción de las especies como grupos aislados reproductivamente (Mayr) y el enfoque filogenético, que las clasifica

atendiendo a la descendencia ancestral común (Cracraft). Se trata de enfoques distintos que dividen a los organismos de formas muy diferentes. Esta variedad de criterios no parece concordar con la pretensión de definir las clases mediante propiedades esenciales.

En el extremo opuesto a este esencialismo respecto a las clases naturales hay quienes defienden un enfoque constructivista, que en su versión más fuerte niega que la naturaleza esté dotada de una estructura que podamos descubrir o representar aproximadamente mediante nuestros conceptos de clase. No hay clases en la naturaleza sino que éstas son simples creaciones arbitrarias nuestras cuya justificación es meramente pragmática. El relativismo ontológico de Goodman (1978, capítulos 6 y 7) constituye una buena muestra de este constructivismo.

Una postura que, aunque distante del esencialismo, no deja de ser realista respecto a las clases naturales consiste en concebir éstas como racimos contingentes pero suficientemente estables de propiedades (p. e. Boyd, 1991, Milikan, 1999). Una clase natural no se caracterizaría por una propiedad esencial, sino por la coexistencia de una serie de propiedades, debida a que una de ellas favorece la presencia de las demás o, a que un mecanismo interno o externo tiende a asegurar la co-ocurrencia de propiedades. Desde este punto de vista, las clases naturales se convierten en una sólida base para inferencias inductivas y para la formulación de leyes de la naturaleza: sólo merecerían este calificativo las proposiciones generales referidas a clases naturales.

Este enfoque da razón de una intuición básica acerca de las clases naturales. Si no consideramos natural la clase de los animales blancos es porque tal concepto no es nada informativo, sólo nos indica el color de los animales que caen bajo a él; y si pensamos que es mucho más natural la clase de los mamíferos, es porque podemos aplicar a sus integrantes todo un racimo de propiedades (respiración pulmonar, desarrollo embrionario dentro de la madre, sangre caliente, etc.). Sin embargo, esta concepción de las clases como racimos de propiedades, lejos de ofrecer una justificación de las leyes, supone ya el concepto de ley. Contar con la persistencia de uno de tales racimos de propiedades equivale a afirmar una conexión nomológica entre ellas.

Pero, por otra parte, intentar encontrar una salida a las discusiones acerca de las clases naturales apelando a la noción de ley significa adentrarse en un territorio donde la diversidad de opiniones es, al menos, tan amplia como en el caso de las clases naturales. Hay, por ejemplo, quienes defienden respecto a las leyes posiciones fuertemente realistas (Armstrong), quienes niegan que propiamente existan leyes (Giere, van Fraassen) y quienes ponen en cuestión su verdad, debido a su carácter abstracto (Cartwright). En lo que sigue voy a esbozar un concepto básico de ley y a deducir algunas consecuencias que se seguirían al construir a partir de él la idea de clase natural.

No es difícil estar de acuerdo en que una ley (de la ciencia o del conocimiento ordinario) es una proposición general (universal o estadística) ampliamente confirmada que cumple una función explicativa o/y predictiva; dos funciones, una de carácter teórico y la otra de carácter pragmático, que, sin embargo, se

contraponen, apuntan en direcciones opuestas. Los principios de las teorías (por ejemplo, el principio de gravitación) gozan de gran capacidad explicativa porque articulan un buen número de leyes de más bajo nivel y otras hipótesis, pero, dado su alto grado de abstracción, no permiten extraer de ellos predicciones precisas; y a la inversa, las leyes más útiles predictivamente, las que se pueden utilizar para obtener predicciones precisas, son leyes con menor capacidad explicativa, como ocurre, por ejemplo, con las leyes que se derivan de principios (por ejemplo, la ley de caída de Galileo).

La derivación de una ley respecto de un principio responde a la incorporación de condiciones iniciales más restrictivas, con lo que la ley pierde nivel de abstracción y, por tanto, generalidad, pero gana precisión en sus predicciones. Sin embargo, como se nos ha recordado durante tanto tiempo, hay una diferencia importante entre las leyes y las que se han llamado generalizaciones accidentalmente verdaderas. Desde el punto de vista que propongo, no se pueden considerar leyes aquellas generalizaciones que, aun siendo nomológicamente explicables, dejan de tener utilidad predictiva debido a su carga de condiciones iniciales, es decir, de premisas. Es cierto que todos los bolígrafos caen al suelo si los soltamos en el aire. Y este hecho general tiene una explicación teórica a partir del principio de gravitación y de una serie de condiciones iniciales relativas a la masa de los bolígrafos, la masa de la tierra, etc., pero no diríamos que se trata de una ley. Y no lo es porque su explicación requiere de un buen número de premisas y no aporta nada predictivamente: no predice nada que la ley de Galileo no prediga mejor. Llegados a este punto podríamos decir que una ley es una proposición general ampliamente confirmada que consigue un aceptable compromiso ente un mínimo de premisas explicativas y un máximo de consecuencias predictivas precisas.

De todos modos, esta versión de las leyes no atiende sólo a la situación de éstas en el marco de una teoría como principios o como leyes de nivel inferior. Dada la importancia pragmática que se concede a su capacidad predictiva, entran en la categoría de ley aquellas generalizaciones bien confirmadas que se muestran útiles predictivamente, aunque carezcan de soporte teórico. Sería el caso de la ley de caída de Galileo antes de su integración en la mecánica de Newton. Esta característica es importante en especial cuando abordamos el carácter nomológico de ciertas regularidades en biología y ciencias sociales y, en general, si tenemos en cuenta que no todo el conocimiento científico está codificado en teorías.

Veamos algunas características de esta concepción básica de las leyes. En primer lugar, como se ha podido ver, no es preciso apelar en ella a la siempre problemática idea de necesidad nomológica. Por otra parte, este concepto de ley es gradual: hay leyes que lo son más que otras, según el balance conseguido entre capacidad explicativa y capacidad predictiva, con lo cual, la frontera entre una ley y una generalización que no merece tal nombre es necesariamente vaga. Por último, esta versión de las leyes es perfectamente compatible con una concepción realista, en el sentido en que la regularidad y las capacidades explicativas y predictivas de las leyes dependen de su conformidad con la estructura nomológica en la naturaleza. Sin embargo, no hay razones para aceptar una correspondencia

directa entre las leyes de la ciencia (o del conocimiento ordinario) y "leyes de la naturaleza". No hay por qué dar por supuesto que la ley de Ohm o la de Coulomb son descubrimientos de las correspondientes leyes naturales. Dado el contraste entre, por una parte, la división de la ciencia en múltiples ramas y el carácter abstracto de sus leyes (no sólo de los principios teóricos), y por otra, la complejidad de los sucesos reales, en los que las leyes de la ciencia sólo se cumplen aproximadamente y en los que normalmente se entremezclan leyes de ciencias muy diversas, parece ingenuo pensar que cada una de las leyes en las que condensamos nuestro conocimiento se corresponde directamente con una ley de la naturaleza.

Si, para finalizar, definimos ahora clase natural como un conjunto de elementos que figura como sujeto de una o varias leyes, estas ideas sobre las leyes permiten extraer algunas breves consecuencias acerca de las clases naturales. En primer lugar, es preciso señalar que, aunque, según la idea de ley que he propuesto, esta definición de clase natural despoja a este concepto de toda connotación de necesidad, no por ello renuncia a una interpretación realista de la clases naturales: la existencia de clases naturales depende de la existencia de leyes que les afecten, y tales leyes dependen en última instancia, como he señalado, de la estructura nomológica de la naturaleza. Pero, a diferencia de lo que defiende el esencialismo, un individuo nunca queda suficientemente definido por su pertenencia a una clase determinada, sino que puede pertenecer del mismo modo a clases diferentes, porque sus propiedades pueden ser objeto de leyes diferentes y de ciencias distintas. Por otra parte, una clase puede ser definida de diferentes modos, según los intereses clasificatorios, atendiendo a distintas regularidades nomológicas. Ya vimos cómo existen en sistemática biológica diferentes conceptos de especie con resultados diferentes en cuanto a los elementos que integran cada grupo. De todo ello se sigue que caben diferentes clasificaciones *igualmente legítimas* para un mismo dominio de la realidad. El mundo es lo suficientemente complejo como para creer que existe un único modo de trincharlo[1].

Referencias bibliográficas

Boyd, R. (1991), "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies* 61, 127–148.

Dupré, J. (1993), *The Disorder of Things*. Cambridge, Mass.: Harvard University Press.

Ellis, B. (2001), *Scientific Essentialism*. Cambridge: Cambridge University Press.

Goodman, N. (1978), *Ways of Worldmaking*. Indianapolis: Hackett.

Millikan, R. G. (1999),"Historical Kinds and the Special Sciences", *Philosophical Studies* 95: 45–65.

[1] Estas conclusiones coinciden con lo que Dupré (1993) defiende como realismo promiscuo o pluralismo ontológico.

Razones y controversias en la experimentación científica

Juan Bautista Bengoetxea
Universidad de Valladolid
bautista@fyl.uva.es

La disputa entre las tesis conceptualistas (epistémicas) y las sociologistas en torno a la experimentación científica se puede situar en un espacio común en la que la confrontación se vuelva complementariedad o reensamblaje (Latour 2005, Nickles 1989). Sin embargo, esta posible complementariedad no significa que todo tipo de razón esté a la par del resto (simetría). Básicamente, cabría afirmar que las razones epistemológicas en la experimentación científica no se pueden eliminar *tout court*. Lo interesante más bien sería procurar articular adecuadamente el juego de las razones (epistemológicas o sociales) (Brandom 1994) en el que se hicieran explícitos los procesos reales de la actividad experimental. Sería cuestión de hacer explícitos los procesos que Nickles etiqueta de «dormidos» o que se han automatizado hasta hacerse transparentes, invisibles. Trato de mostrar esto en el caso del *problema del neutrino solar*.

Razones

La filosofía de la ciencia y de la tecnología interesada en la experimentación científica no parte abiertamente de ningún enfoque aceptado de la racionalidad. No obstante, a menudo se da por supuesto que el discurso filosófico que recurre a razones se sostiene sobre la lógica, la metodología o alguna estrategia epistemológica tradicional. Pero esto no tiene por qué ser así. Las razones pueden ser de índole diversa en el desarrollo y avance de la ciencia, tal y como las concepciones de la *racionabilidad* de Toulmin (2001) y del *razonamiento práctico* de Kovac (2002), entre otras, mantienen.

Los trabajos de Allan Franklin (1990, 2002) y de Trevor Pinch (1985) son aquí ejemplo de dos formas opuestas de entender el uso de las razones en la ciencia experimental. Franklin defiende la validez de las estrategias epistemológicas y Pinch la niega con el objetivo de sustituirla radicalmente por el 'razonamiento' sociológico. Lo que para uno son razones, para el otro son sinrazones. Ciertamente, ambas aportaciones han sido interesantes para la filosofía de la experimentación científica, pero ni son suficientes ni se ajustan completamente a la práctica científica real. La concepción de *dar y pedir razones*, insertada en la tarea de hacer explícito el discurso práctico-inferencial, es una forma conveniente de comprender más adecuadamente la experimentación científica y el papel que ésta desempeña en la elección de teorías. Mi punto de partida es el siguiente:

(i) Observo un caso de la práctica científica real porque los enfoques normativistas no satisfacen por lo general el *requisito de realizabilidad* (Nickles 1989: 317s), y

(ii) Critico la visión *esencialista* de las propuestas de Pinch y de Franklin. Ambas colapsan el debate acerca de la experimentación, como si el resultado de ésta fuera algo inmóvil y unidireccional una vez aceptado en las fases iniciales de la investigación. No es así. La investigación experimental perdura y se extiende a lo largo de un proceso de construcción y reconstrucción constante en el que cabe toda una variedad de recursos cognitivos y no cognitivos para justificarlos, sean epistémicos o sean sociales.

Controversias

Parte Franklin de una idea en la que ha insistido desde hace dos décadas: la ciencia es una empresa razonable que se basa en pruebas empíricas experimentales (evidencias), en la crítica y en la discusión racional. Propone una epistemología del experimento consistente en un conjunto de estrategias que suministran *creencia razonable* en los resultados.

Los elementos más destacables que su propia estrategia epistemológica procura satisfacer son éstos (Franklin 2002: 3-5; 1990: 104ss): (1) los chequeos y la calibración; (2) la reproducción de artefactos, (3) la eliminación de fuentes plausibles de error, (4) el uso de los resultados para argumentar a favor de su propia validez, (5) el uso de teorías de los fenómenos que estén bien corroboradas independientemente, (6) el uso de un aparato basado en una teoría bien corroborada, y (7) el uso de argumentos estadísticos. El objetivo de estas estrategias es generar creencias razonables en la validez de un resultado experimental. Básicamente, permiten distinguir entre una observación válida (o medición válida) y un artefacto (que el aparato experimental pueda hacer surgir).

Pinch, Collins y Pickering, entre otros, han dado lugar a objeciones al enfoque de que los resultados experimentales se aceptan sobre la base de argumentos epistemológicos. MacKenzie (1989: 412) lo remarca: los sociólogos de la ciencia han mostrado que ningún experimento puede por sí mismo *conducir a* la solución a una controversia o dar razón a la aceptación de un hecho concreto. Cualquier crítico convencido podría encontrar una razón para oponerse a un supuesto resultado.

Según Franklin, MacKenzie está originando dudas acerca de la validez de los experimentos a la hora de contrastar teorías o hipótesis. Pinch lo expuso claramente en su propuesta de una sociología del experimento.[1] Pinch (1985: 175s) aplica su interpretación sociologista al caso de la contrastación de la teoría astrofísica nuclear por medio de los experimentos de detección de neutrinos solares. Interpreta de una manera singular los resultados de Davis y Bahcall (Cf. Davis & Bahcall 1982). Destacaré las siguientes dos notas de su interpretación:

[1] Aunque extensible a toda lógica pretendida de la investigación científica, su crítica en este artículo está dirigida en particular al falsacionismo de Popper.

Las *discrepancias*: respecto de los resultados de Davis, Pinch (1987: 177) apunta a la existencia de *discordancias* entre éstos y los de otros experimentadores, debidas ante todo a la falta de replicaciones del experimento. Parece que los elementos económicos (sociales) se impusieron.

Invalidez epistemológica y *validez social* en la toma de decisiones: según Pinch, Bahcall no compartía el criterio de Davis –según el cual las observaciones refutaban la teoría—. Bahcall no estaba convencido de que hubiera contradicción entre la teoría y el experimento y continuó experimentado con el fin de satisfacer, según Pinch, prioridades no epistémicas (prestigio, apoyos institucionales, etc.). Por consiguiente, las pruebas empíricas parecían no contar con el apoyo epistemológico suficiente a la hora de decidir si tomarlas en consideración o no.

2001, el problema del neutrino solar

La cuestión: la hipótesis básica barajada por los astrofísicos hasta 1964 era la siguiente:

$$(H)\ 4p \rightarrow {}^{4}He + 2\ e^{+} + 2\nu_e$$

Cuatro núcleos de hidrógeno, los protones (p), se transforman en el interior solar en núcleos de helio (^{4}He), en dos positrones (e^+) y en dos neutrinos (ν_e). Según el modelo estándar de la física de partículas, los neutrinos –sean electrónicos (ν_e), muónicos (ν_μ) o tauónicos (ν_τ)— tienen carga eléctrica nula, apenas interactúan con la materia y, ante todo, no poseen masa.

Con el fin de contrastar H, en 1964 Raymond Davis y John Bahcall diseñaron un experimento de «generación de energía nuclear en las estrellas» (Bahcall 1964: 300). En su vertiente teórica, Bahcall realizó los cálculos de la cantidad de neutrinos de energías diferentes producidos por el sol según un modelo computacional detallado del sol. Con el mismo modelo calculó a su vez la cantidad de átomos de argón radioactivo (^{37}Ar) que los neutrinos solares producirían en el experimento que Davis llevaría a cabo en un tanque de percloroetileno (C_2Cl_4) (Davis 1964: 303).

Davis obtuvo sus primeros resultados en 1968 y en ellos se detecta únicamente 1/3 de la cantidad de átomos de argón radioactivo predichos. La discordancia entre la predicción y el resultado de la medición era evidente, pero los científicos no consideraron el abandono de H ni de la teoría astrofísica nuclear. Al contrario, y con el objetivo de continuar indagando en el tema, la comunidad científica bautizó el caso como *el problema del neutrino solar*.

La validez social en la toma de decisiones: Pinch plasma una interpretación únicamente sociológica (1985: 177ss) de la labor de Bahcall con el fin de poner en duda la validez de los criterios lógicos popperianos en el caso de la contrastación de la teoría astrofísica. Los datos que utiliza a favor de su interpretación se refieren casi exclusivamente a los datos de entrada iniciales que Bahcall manejó continuamente con el fin de adecuar los resultados teóricos astrofísicos a los experimentales. Reconoce que Bahcall se esforzó en hacer cada vez más fino el experimento, pero enfatiza que ninguna de sus mejoras intentó revisar la teoría

básica, sino tan solo los parámetros de entrada necesarios para el cálculo de los ajustes.

El juego de las razones mediante explicitación: en una de sus formas, al menos, consistiría en hacer explícitos procedimientos varios que a menudo se hacen invisibles o transparentes en los procesos de desarrollo y reconstrucción experimental. Los procedimientos se pueden catalogar en virtud de los diversos experimentos realizados, y en todos ellos es indispensable la puesta a punto de mecanismos tales como los ajustes de mediciones y los exámenes continuados de cálculos y modelos teóricos.

En el caso del neutrino solar, los grupos de experimentos realizados son los siguientes: el experimento Kamiokande (1989) (detector de agua pura en vez de un detector de cloro, como lo era el original), los experimentos GALLEX y SAGE (1990s) (varios detectores de galio), el experimento Super-Kamiokande (1990s) (mediciones más precisas de neutrinos de altas energías) y los experimentos del Sudbury Solar Neutrino Observatory (SNO) (2001) (primeros resultados sobre el neutrino solar obtenidos con un detector de 1000 toneladas de agua pesada (D_2O)). Si el modelo estándar de la física de partículas fuera correcto, la fracción medida por el SNO y la fracción medida por el Super-Kamiokande deberían ser la misma. Es decir, todos los neutrinos deberían ser neutrinos electrónicos. Pero las fracciones eran distintas (1/3 y 1/2 respectivamente), por lo que el modelo estándar de la física de partículas era erróneo.

Combinando las mediciones del SNO y del Kamiokande, el grupo SNO determinó la cantidad total de neutrinos solares de todos los tipos, así como la cantidad de únicamente los neutrinos electrónicos. La cantidad total concordaba con la cantidad predicha por el modelo computacional del sol. Los neutrinos electrónicos constituyen aproximadamente 1/3 de la cantidad total de neutrinos.

La prueba clave fue descubierta. Se trataba de la diferencia entre la cantidad total de neutrinos y la cantidad de neutrinos electrónicos. Los neutrinos desaparecidos en realidad estaban presentes, pero en una forma mucho más difícil de ser detectada. Era la forma de los neutrinos muónicos y tauónicos.

Los nuevos experimentos, vinculados directamente a innovaciones tecnológicas, son muestra del uso y mejora de procesos razonados conceptualmente, pertenecientes también a contextos sociales que influyen en ellos, y posibles causantes de cambios de teoría. El problema del neutrino solar, aún parcialmente abierto, recibió una primera solución satisfactoria en 2001, no en 1985. La interpretación sociologista que negaba las razones epistémicas fue precipitada.

Referencias bibliográficas

Bahcall, J. N. (1964), 'Solar Neutrinos, II: Theoretical', *Physical Review Letters* 12 (11): 300-302.

Bahcall, J. N. y Davis, R. (1982), 'An Account of the Development of the Solar Neutrino Problem', en C. A. Barnes, D. D. Clayton y D. Schramm, eds. *Essays in Nuclear Astrophysics*, Cambridge, Cambridge University Press, pp. 243-285.

Brandom, R. (1994), *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, Harvard University Press.

Davis Jr., R. (1964) 'Solar Neutrinos, II: Experimental', *Physical Review Letters* 12 (11), pp. 303-305.

Franklin, A. (1990), *Experiment: Right or Wrong*, Cambridge, Cambridge University Press.

— (2002), *Selectivity and Discord: Two Problems of Experiment*, Pittsburgh, Pa., University of Pittsburgh Press.

Kovac, J. (2002), 'Theoretical and Practical Reasoning in Chemistry', *Foundations of Chemistry* 4/2, pp. 163-171.

Latour, B. (2005), *Reensamblar lo social: Una introducción a la teoría del actor-red*, Buenos Aires, Manantial, 2008.

Mackenzie, D. (1989), 'From Kwajalein to Armageddon? Testing and the social construction of missile accuracy', en D. Gooding, T. Pinch y S. Schaffer (eds.), *The Uses of Experiment: Studies in the Natural Sciences*, Cambridge, Cambridge University Press, pp. 409-435.

Nickles, T. (1989), 'Justification and Experiment', en D. Gooding, T. Pinch y S. Schaffer (eds.), *The Uses of Experiment: Studies in the Natural Sciences*, Cambridge, Cambridge University Press, pp. 299-333.

Pinch, T. (1985), 'Theory Testing in Science — The Case of Solar Neutrinos: Do Crucial Experiments Test Theories or Theorists?', *Philosophy of the Social Sciences* 15/2, pp. 167-187.

Toulmin, S. (2001), *Regreso a la razón: el debate entre la racionalidad y la experiencia y la práctica personales en el mundo contemporáneo*, Barcelona, Península, 2003.

Algunos problemas en torno a la evaluación del éxito teórico

María de la Concepción Caamaño Alegre
Universidad de Valladolid
mariac@fyl.uva.es

Una gran parte de las teorías científicas del pasado, como el geocentrismo, la teoría del flogisto o la mecánica newtoniana, han sido falsadas. Por lo que, si sólo aplicamos la verdad como criterio de éxito de una teoría, a ninguna de ellas se le podría reconocer ningún éxito teórico. Por otra parte, el recurso a nociones (veritativas) graduales, como las de verosimilitud o aproximación a la verdad, no resulta aplicable satisfactoriamente en la comparación de teorías debido a diversos problemas, como son la infradeterminación de la teoría por los hechos, la independencia entre la verdad de una teoría y su éxito predictivo, o el problema de la inconmensurabilidad de los lenguajes teóricos.

El punto de partida del análisis que aquí se desarrolla es, precisamente, el reconocimiento de los distintos problemas que plantea la inaplicabilidad de criterios veritativos como los únicos o principales en la evaluación del éxito teórico. El propósito del presente trabajo es el de realizar un análisis crítico de los distintos factores que se suelen considerar determinantes del éxito teórico. Se llamará la atención sobre los errores más comunes relativos al establecimiento de dichos factores, que, en general, pueden deberse, tanto a la insuficiencia o inaplicabilidad de éstos para la evaluación del éxito teórico, como a la falta de reconocimiento de otros que sí resultan esenciales para tal evaluación. Entre los primeros habría que subrayar, además del factor veritativo ya apuntado, el del éxito empírico independiente de la teoría, así como el de la adecuación empírica entendida como mera corrección observacional de una teoría. Entre los segundos, esto es, entre los factores habitualmente obviados, conviene destacar, por un lado, la adecuación empírica concebida en un sentido más pragmáticamente comprometido, como éxito predictivo comparativamente superior al de otras teorías, y, por otro lado, el éxito explicativo en todas sus demás vertientes (simplicidad, unificación, analogía). El estudio de los distintos factores mencionados permitirá entender mejor en qué sentido el éxito de la teoría del flogisto, o la del calórico, fue mucho más limitado y actualmente estéril que el de teorías como el geocentrismo o la mecánica newtoniana, a pesar de ser todas ellas teorías falsadas, y no susceptibles de comparación en términos de verosimilitud con sus respectivas teorías sucesoras.

Factores inaplicables o insuficientes para la determinación del éxito teórico

Un primer factor que se revela inaplicable para evaluar el éxito de una teoría viene dado por el progreso empírico que suele asociarse con ella, pero que, en realidad,

es independiente del desarrollo de la misma. Ha de aclararse que el éxito teórico se entenderá aquí como aquel éxito debido a una teoría, y no como el que meramente acompaña a su desarrollo. La anterior aclaración nos sitúa ya en un punto importante del debate en torno a esta noción, pues, tras la advertencia de I. Hacking (1983) acerca del alto grado de independencia del desarrollo experimental con respecto al teórico, autores realistas como S. Psillos (1994), P. Kitcher o J. Worrall han seguido interpretando que teorías falsadas, como la del calórico, supusieron un gran avance teórico preservado por teorías sucesoras, mientras que autores instrumentalistas como L. Laudan o H. S. Chang (2003) han entendido que los únicos componentes preservados en ese caso han sido independientes de la teoría. Recordemos brevemente que, desde la teoría del calórico, se postulaba la existencia de un fluido sin peso y auto-repelente llamado 'calórico', que se expandiría desde cuerpos más calientes a cuerpos más fríos. Ciertamente, durante el desarrollo de esta teoría se realizaron muy importantes avances experimentales en el estudio del calor, como por ejemplo, la determinación del coeficiente de expansión conforme a la temperatura y según el estado de agregación de la materia, o la determinación del calor específico de distintas sustancias. Sin embargo, incluso si la mayoría de los fenómenos relacionados con el calor resultaban interpretables como procesos causados por la acción del calórico, las propiedades involucradas de manera esencial en las predicciones cuantitativas exitosas sólo eran relativas a la temperatura. Por otra parte, las abundantes predicciones cualitativas, como aquellas concernientes a la transmisión del calor, eran sugeridas simplemente por las observaciones ordinarias. En otras palabras, la postulación del calórico no era necesaria, ni siquiera útil, para inferir muchas de las consecuencias empíricas corroboradas acerca del calor, que dependían, en cambio, de leyes y generalizaciones empíricas relativas a la temperatura.

Un segundo factor a comentar es el de la verdad como criterio de éxito teórico. Como se apuntó al comienzo, dicho factor, al contrario de lo que se suele defender desde enfoques realistas, se muestra insuficiente para poder determinar cuidadosamente el éxito teórico relativo de cada teoría, es decir, la superioridad de una teoría en comparación con otras rivales. Incluso autores agnósticos con respecto a la tesis de la inconmensurabilidad, como P. Lipton (cfr. 1991/2004, pp. 57, 69, 70) o R. M. Forster (cfr. 2002, p. 233), reconocen que problemas como la infradeterminación de la teoría por los hechos, o la independencia entre verdad teórica y éxito predictivo (que en cambio sí dependería de asunciones falsas en forma de idealizaciones), nunca nos permitirían establecer el grado de aproximación a la verdad de una teoría de manera concluyente. En algunos casos, incluso, las teorías falsas, dado su éxito predictivo, parecerían aproximarse más a la verdad que teorías verdaderas menos rentables desde el punto de vista predictivo.

Un último factor, de índole más empiricista, pero igualmente insuficiente para la evaluación del éxito teórico, es la adecuación empírica de una teoría en su acepción estrictamente consecuencialista de corrección observacional. En un interesante análisis de la expresión '*saving the phenomena*', L. Laudan (cfr. 2002,

pp. 168-9) distingue tres sentidos fundamentales en los que ha sido usada dentro del contexto filosófico: como corrección observacional (*á la* van Fraassen), implicando, por tanto, que todas las consecuencias observacionales de una teoría son verdaderas, como explicación de todos los hechos importantes conocidos dentro de un determinado dominio de investigación (según la concepción de Platón), y, finalmente, como corrección observacional más capacidad explicativa para dar cuenta de todos los fenómenos ya explicados por teorías precedentes. Laudan argumenta convincentemente que, si entendemos la adecuación empírica exclusivamente en el primer sentido (el 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.

Veamos ahora cuáles son los factores que algunos de los autores ya mencionados proponen como los determinantes para la evaluación del éxito teórico.

Factores relevantes habitualmente ignorados en la determinación del éxito teórico

Uno de los factores decisivos es el de la adecuación empírica entendida en la tercera de las acepciones distinguidas por Laudan (es decir, como corrección observacional más capacidad explicativa para dar cuenta de todos los fenómenos ya explicados por teorías precedentes), unido a la aportación de recursos predictivos para nuevos fenómenos (cfr. 2002, p. 166). Conviene señalar que, desde el enfoque estructuralista, se ha desarrollado una noción de progreso científico en esta misma clave. Así, según la caracterización de progreso científico ofrecida por U. Moulines (2000), aquél debería concebirse, no como una acumulación de verdades sino como una acumulación de problemas resueltos, es decir, como un saber más sobre algunos subdominios idénticos u homomorfos de aplicaciones intencionales paradigmáticas de especial relevancia epistemológica o pragmática. Desde su enfoque, la posibilidad de extender más aplicaciones intencionales, o, simplemente, aplicaciones más interesantes, a los modelos de una teoría equivale a poder resolver más problemas (epistémico-pragmáticos) o problemas más interesantes. Volviendo al tratamiento de Laudan sobre este punto, cabe señalar, como una de sus conclusiones más importantes, la necesidad de superar el modelo consecuencialista de relevancia evidencial, conforme al cual, la evidencia relevante para una teoría se halla determinada por las consecuencias derivables de dichas teoría. Como ejemplo de evidencia relevante para una teoría, y a la vez no determinada por sus consecuencias, menciona lo que él denomina 'las anomalías no refutatorias', esto es, los fenómenos que resultan explicables por otras teorías distintas de la que se evalúa, sin que, por otra parte, dicho fenómenos la refuten. Lo que pondrían de relieve este tipo de factores sería la importancia de la completitud o el alcance explicativos de una teoría frente a la mera corrección observacional. Abundando en esta idea Laudan apunta que, 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, tendríamos que conformarnos

con modestas generalizaciones empíricas o afirmaciones sobre hechos particulares renunciando a la teorización (cfr. 2002, pp. 170-2).

Todavía desde un punto de vista consecuencialista, pero tratando de afinar los recursos para evaluar la calidad del apoyo evidencial recibido por una teoría, Lipton y Forster son algunos de los autores que mantienen la superioridad epistémica de la capacidad predictiva frente a la capacidad para acomodar los fenómenos, indicando así otro factor clave para la evaluación del éxito teórico. En efecto, la distinción de Forster entre contrastación diacrónica y sincrónica (cfr. 2002, p. 233) coincide con la de Lipton entre predicción y acomodación (cfr. 1991/2004, p. 68). La contrastación diacrónica, al igual que la predicción, requeriría que las hipótesis se contrastasen a partir de un conjunto de datos diferentes de aquellos empleados para construir la teoría. La contrastación sincrónica, en cambio, consistiría en la acomodación de los datos que se han empleado para elaborar la hipótesis. La superioridad de la predicción sobre la acomodación se debería a que, en la segunda, el éxito de la hipótesis para ajustarse a la evidencia sería atribuible a que la hipótesis había sido diseñada precisamente con ese fin, sin reflejar la estructura real de los fenómenos y, en consecuencia, sin poder aplicarse a fenómenos nuevos del mismo tipo.

Un último factor a destacar es el de las posibilidades de comprensión de los fenómenos abiertas por una teoría, en otras palabras, lo que Lipton denomina '*loveliness*' en relación con una explicación (cfr. 1991/2004, p. 59). A diferencia del factor '*likeliness*', determinado por el éxito predictivo, el factor '*loveliness*' sería relativo a la capacidad de una explicación para simplificar, unificar y someter a analogías familiares aquellos fenómenos que intentan explicarse. Quizá resulte de interés reproducir, en este punto, la distinción de P. Thagard (1978) entre tres criterios explicativos para la elección entre teorías, a saber, los de "*consilience*" (poder unificador y sistematizador de una teoría), simplicidad (uso limitado de hipótesis auxiliares), y analogía (similitud entre modelos teóricos consolidados y nuevos modelos teóricos). Cada uno de ellos aporta claves valiosas para poder explorar las distintas formas de avance teórico detectables en el desarrollo de la ciencia.

Si atendemos a ejemplos históricos centrándonos en las distintas posibilidades de éxito teórico apuntadas más arriba, comprobaremos que la contribución proveniente de teorías como las del flogisto o del calórico es notablemente distinta de aquella procedente de teorías como el geocentrismo o la mecánica newtoniana, donde los modelos teóricos han conservado no sólo una enorme utilidad como instrumentos de cálculo predictivo, sino también un importante valor explicativo.

Conclusión

El análisis propuesto en este trabajo permite diferenciar cuatro casos distintos, no excluyentes, de éxito teórico: aquel derivado de la mayor adecuación empírica de una teoría frente a otra, en el sentido de Laudan (teoría del oxígeno frente a teoría del flogisto), el relativo a la superioridad de la predicción frente a acomodación (teoría de la energía cinética del calor frente a teoría del calórico), el que se

desprende de la superioridad sistematizadora y unificadora de la teoría (mecánica newtoniana frente a mecánica galileana), y el propiciado por una relación de analogía con respecto a teorías consolidadas (teoría ondulatoria frente a teoría corpuscular de la luz).

Referencias bibliográficas

Chang, H. S. (2003) "Preservative Realism and its Discontents: Revisiting Caloric", *Philosophy of Science* 70 (5), pp. 902-912.

Forster, M. R. (2002) "Hard Problems in the Philosophy of Science: Idealisation and Commensurability", en R. Nola y H. Sankey (eds.), *After Popper, Kuhn and Feyerabend. Recent Issues in Theories of Scientific Method*, Dordrecht, Kluwer Academic Publishers, pp. 231-250.

Hacking, I, (1983) *Representing and Intervening. Introductory Topics in the Philosophy of Natural Science*, Cambridge, Cambridge University Press.

Laudan, L. (2002) "Is Epistemology Adequate to the Task of Rational Theory Evaluation?", en R. Nola y H. Sankey (eds.), *After Popper, Kuhn and Feyerabend. Recent Issues in Theories of Scientific Method*, Dordrecht, Kluwer Academic Publishers, pp. 165-175.

Lipton, P. (1991) *Inference to the Best Explanation*, New York, Routledge, 2004.

Moulines, C. U. (2000) "Is There Genuinely Scientific Progress?", en A. Jonkisz y L. Koj (eds.), *On Comparing and Evaluating Scientific Theories, Poznan Studies in the Philosophy of the Sciences and the Humanities*, Amsterdam, Rodopi, pp. 173-197.

Psillos, S. (1994) "A Philosophical Study of the Transition from the Caloric Theory of Heat to Thermodynamics: Resisting the Pessimistic Meta-Induction", *Studies in the History and Philosophy of Science* 25, pp. 159-190.

Sankey, H. (2002) "Methodological Pluralism, Normative Naturalism and the Realist Aim of Science", en R. Nola y H. Sankey (eds.), *After Popper, Kuhn and Feyerabend. Recent Issues in Theories of Scientific Method*, Dordrecht, Kluwer Academic Publishers, pp. 211-229.

Thagard, P. R. (1978) "The Best Explanation: Criterion for Theory Choice", *The Journal of Philosophy* 75 (2), pp. 76-92.

Mathematical Indispensability*

Eduardo Castro
University of Beira Interior and LanCog
Philosophy Centre of the University of Lisbon
ecastro@ubi.pt

The Argument of Quine-Putnam

Argument Q-P

- ($1_{Q\text{-}P}$) Our best scientific theory of the world makes indispensable use of mathematical things. (This is taken as an unvarnished fact.)
- ($2_{Q\text{-}P}$) To draw a testable consequence from our theory requires the use of various far-flung parts of that theory, including much mathematics. (This is confirmational holism.)
- ($3_{Q\text{-}P}$) Our theory is committed to those things that it says 'there are'. (This is Quine's criterion of ontological commitment.)
- ($\therefore$) Our theory, and we who adopt it, are committed to the existence of mathematical things. (Maddy 1997, p. 133)

Maddy's Objections

Maddy attacks the Quinean view about natural science that supports the argument Q-P, namely, the doctrine of confirmational holism and the criterion of ontological commitment.

First objection

The first objection aims to show that the alleged indispensability of empirical entities in our best scientific theories is not a sufficient condition to convince (all) the scientists that these entities exist. Maddy argues that, at the end of the nineteenth century, there was a discrepancy between scientific practice and Quine's epistemological view around atomic theory (Quine 1960). This theory was endowed with all the five Quinean theoretical benefits (simplicity, familiarity of principle, scope, fecundity, empirical confirmation) but the scientists, 'as a whole', did not accept the atomic theory as true. This shows that "scientists do not, in practice, view the overall empirical success of a theory as confirming all its parts" (Maddy 1997, p. 142); scientific hypotheses are taken as useful fictions until further and more conclusive tests were carried on.

* Research supported by FCT, grant nº SFRH/BPD/46847/2008.

Second Objection

Maddy's second objection is a mere extension of the first objection to the mathematical domain: if the ontological commitment to some scientific entities (posited in our best scientific theories) is disputable, the ontological commitment to some mathematical entities (posited in our best scientific theories) is also disputable.

Based on an analysis of scientific theories she concludes that the alleged continuity of physical space-time is an open question and that we have no *indispensability* reasons to commit to the mathematical large finite (*a fortiori*, we have no reasons to commit to the mathematical infinity), since our best theories say that the numbers of particles of the Universe is finite.

Replies to Maddy's objections

Reply to first objection

Firstly, there are alternative histories about atomic theory, in particular, about kinetic theory, under which Maddy's history is subsumed, which are consistent with Quine's epistemology. Secondly, new versions of the indispensability argument, that emerge from a slight change of one of the five benefits –empirical confirmation–, accommodate Maddy's objection.

According to Stephen Brush (1968 and 1974), kinetic theory was empirically (and metaphysically) confirmed before Perrin's experiments, and all disbelief about the truth of kinetic theory among scientists was ideologically rooted. Those scientists can be seen like the anti-realists philosophers of nowadays (or from the past) that deny the existence of unobservable entities by philosophical or ideological reasons. In other words, first philosophy was practised by some scientists from the late nineteenth century. However, almost all kinds of naturalism repudiate first philosophy. Being a naturalist is not just looking at scientific practice and describing it. Inside of naturalism there is room for prescription, advice and recommendation. Overall, we should not focus on the scientists' opinions, but analyse directly the scientific theories and experiments. Naturalists should take that type of scientists' opinions against the existence of atoms and the truth of kinetic theory (before and after Perrin's experiments) as psychologically relevant but scientifically inconsequential. Primarily, the analysis and change of ideological scientists' opinions in time is a matter for the sociology of science and not a matter for philosophy.

According to Alan Chalmers (2009), before Perrin's experiments, kinetic theory had some recalcitrant empirical problems and atoms could be bracketed off from kinetic theory, without loss of empirical content. After Perrin's experiments, some empirical problems were solved and there is an empirical content that implied the existence of atoms. So according to Chalmer's view, before Perrin's experiments, kinetic theory was not endowed with the five benefits. The kinetic theory was not endowed (1) with the confirmational benefit (other theories, such

as, thermodynamics, are better theories than kinetic theory) and (2) with the simplicity benefit (if two theories have the same empirical content, then, according to the five benefits, we must choose the simplest theory, in particular, the one with less ontology).

Maddy's historical analysis is subsumed under the histories of Brush and Chalmers. She acknowledges that, around 1900, there were scientists that denied the existence of atoms based on ideological considerations (Brush), but there were also scientists that denied the existence of atoms based on reasonable scientific considerations (Chalmers). That is, the set of scientists that denied the existence of atoms were of two sorts; motivated by ideological reasons (Brush) and motivated by scientific reasons (Chalmers). Now, we are faced with a sort of argument by cases. If Brush's account for the denying of atoms is correct, then kinetic theory must have been considered a true theory around 1900, since the truth value of a theory is not given by first philosophy's views. If Chalmers' account for the denying of atoms is correct, then kinetic theory was false (and not endowed with the five benefits) around 1900, because there were better theories than kinetic theory with the atoms' assumption. In both cases, Quine's epistemology agrees with the state of kinetic theory around 1900. Thus, Maddy's conclusions against Quine's epistemology do not seem to follow from her historical analysis.

Now the second point of the reply. Maddy's first objection is not an objection to confirmational holism or to Quine's criterion of ontological commitment, but an objection to what counts for the scientific value of a scientific theory, since the doctrine of confirmational holism and the criterion for ontological commitments are independent of the epistemic standards that scientific theories must obey. In particular, arguing that atomic theory was endowed with Quine's five theoretical benefits and that atomic theory was not accepted as true by scientists, does not imply that either confirmational holism or Quine's criterion of ontological commitment are incorrect. Thus, a rescue of the thesis that mathematics are indispensable to natural science is achieved by just changing Quine's five theoretical benefits. In this case, a refinement of one of the five benefits – empirical confirmation – will be enough. Here is an example of that refinement:

Argument Q-P (Mild)

($1_{\text{Q-Pmild}}$) Scientific entities are indispensable to scientific theories that are not our current best scientific theories and are waiting for better confirmational tests.

($2_{\text{Q-Pmild}}$) Our scientific theories, which are waiting for better confirmational tests, are empirically tested as a whole.

($3_{\text{Q-Pmild}}$) We should be moderately committed to all and only those entities that are indispensable to scientific theories, which are waiting for better confirmational tests. That is, we should be moderately committed to those entities that

	are quantified in scientific theories, which are waiting for better confirmational tests.
(∴)	We should be moderately committed to scientific entities that are not our current best scientific theories and are waiting for better confirmational tests.

Reply to second objection

My reply is to advance counter-examples to the claim that we do not have indispensability reasons to commit to the mathematical infinity or to the mathematical continuum.

Currently, the Universe is in a state of expansion. In light of the indeterminate values of the Hubble parameter and of the total energy density (of the actual Universe) three cosmological hypotheses of evolution are possible. The first hypothesis says that there will be a 'big crunch' (the Universe will collapse): in this scenario, space and time are finite. The second hypothesis says that the Universe will expand in an asymptotic way (i.e. the expansion of the Universe will never exceed a given limit): in this scenario, space is finite *but* time is infinite. The third hypothesis claims that the Universe will expand forever (the Universe is open): in this case, space and time are infinite. Maddy considers the third hypothesis (the most probable hypothesis) as an open hypothesis and, so, according to her, there are no indispensability reasons to commit ourselves to the mathematical infinity (Maddy 2007, p. 328). However, according to the standard cosmological model, for $t = 0$ there is a singularity; the origin of the Universe (including the beginning of space-time) is a singular state characterised by infinite values of density, temperature and curvature (of space-time). Thus, we have indispensability reasons to commit to the mathematical infinity, since this is an entity indispensable in the standard cosmological model of the Universe for $t = 0$.

The doubts emerged by Maddy, about the continuity of physical space-time, are rooted at micro-level where quantum physics is the best theory for the explanation of the phenomena. However, the physical space-time is not assumed as discrete for scientific theories of quantum mechanics. For example, the standard quantum theory does not imply that there is a discrete nature of space, time or energy. The discrete numbers (energy, angular momentum, etc.) obtained in quantum mechanics came from the compactness of the space where the analysis is carried out. For example, in a non-compact space the solutions of $\nabla^2\phi = -k\phi$ (the typical form of the equation of a harmonic system) are unrestricted eigenvalues for k. Moreover, the possible results of our measurements (the eigenvalues) are represented by real numbers. Why does this happen?

The mathematical structure of Hilbert space is one of paradigms for quantum theory. This space is modelled by real numbers. So, the use of real numbers in quantum theory seems to follow on from the assumption that Hilbert space should be modelled by real numbers, i.e., the continuum.

As far I know, the only doubts concerning the applicability of real numbers arises at the 'Planck scale' (10^{-35} metres), where Quantum Field Theory (the theory emerged by the unification of Special Relativity with Quantum Mechanics) tries to be unified with gravity described by General Relativity. The contemporary research programmes are known as 'Quantum Gravity' and this is an example of a theory focused on in Maddy's discussion.

General Relativity and Quantum Field Theory both consider that space and time are aspects of a space-time represented by a *continuous* differential manifold of four dimensions (a Minkowskian space-time), with a Lorentzian metric associated to this manifold. For Quantum Field Theory this metric is fixed while for General Relativity this metric is dynamical. The main problem of Quantum Gravity is to conciliate these two constraints about space-time, and this is a conceptual problem, since currently there is no empirical problem that could not be solved by current scientific theories. The motivation for Quantum Theory is, for example, the naturalness emerged by combining the three fundamental constants, the principle of unification between fundamental theories or, simply, do not accept the space-time singularities implied by General Relativity. In the last forty years, several theoretical proposals were advanced. However, these proposals are completely absent of empirical confirmation, that is, there is no recognized experimental evidence of the predictable effects by those research programmes.

References

Brush, S. (1968), 'A history of random processes, I. Brownian movement from Brown to Perrin.' *Archive for History of Exact Sciences*, 5, pp. 1-36.

Brush, S. (1974), 'Should the history of science be rated x?' *Science*, 183, pp. 1164-1172.

Chalmers, A. (2009), *The scientist's atom and philosopher's stone –how science succeeded and philosophy failed to gain knowledge of atoms*, USA, Springer-Verlag.

Maddy, P. (1997), *Naturalism in mathematics*, New York, Oxford University Press.

Maddy, P. (2007), *Second philosophy – a naturalistic method*, Oxford, Oxford University Press.

Quine, W. (1960), 'Posits and reality', in his *The Ways of Paradox and other essays*, New York, Random House, pp. 233-241.

Algunas influencias del racionalismo crítico en el anarquismo epistemológico de Feyerabend

Esteban Céspedes
Pontificia Universidad Católica de Valparaíso
estebancespedes@aol.com

1. Una distinción ampliamente aceptada en el ámbito de la filosofía de la ciencia es la que hay entre el racionalismo crítico y el anarquismo metodológico. La diferencia es clara: mientras que el racionalismo crítico es un criterio metodológico de demarcación y de racionalidad, el anarquismo epistemológico afirma que no existen reglas metodológicas únicas que gobiernen el progreso del conocimiento. Esta última perspectiva se resume y es conocida a través del principio propuesto por Feyerabend del "todo vale". Al mismo tiempo, podríamos resumir el racionalismo crítico mediante un principio que dijera "si no es refutable, no vale".

Muchas veces se alude a esta distinción como si ambas tendencias epistemológicas fuesen extremadamente contrarias. Incluso suele identificarse el anarquismo metodológico con el relativismo científico [Preston (1997), p. 7], como una respuesta en contra del racionalismo crítico y de toda concepción filosófica que sostenga que la razón científica garantiza el progreso del conocimiento. Después de todo, el mismo Feyerabend reprocha enérgicamente al racionalismo crítico en su *Tratado contra el método*[1].

Otra clara diferencia entre ambas doctrinas es la manera de ver la práctica científica en general. Mientras que Feyerabend era considerado por la revista *Nature* como "el peor enemigo de la ciencia" [Theocharis y Psimopoulos (1987), pp. 595 – 598], Popper siempre afirmó que la ciencia era la mejor expresión del conocimiento humano[2].

El anarquismo epistemológico se opone, en términos generales, a las metodologías que promueven los criterios de demarcación, entre las cuales se encuentra el racionalismo crítico. Sin embargo, en este trabajo se pretenderá mostrar que existe una relación de condicionalidad entre estos dos sistemas. Esta relación de condicionalidad se puede manifestar en dos sentidos. En un primer sentido, el anarquismo epistemológico es simplemente una consecuencia posible del racionalismo crítico, mientras que en un segundo sentido, es una consecuencia necesaria del mismo. Según esta última interpretación, el anarquismo

[1] "¿Es posible tener ambas cosas, una ciencia tal y como la conocemos y las reglas de un racionalismo crítico como lo acabamos de describir? La respuesta a esta pregunta parece ser un firme y resonante NO" [Feyerabend (1975), p. 162].
[2] "Estoy convencido de que el estudio del desarrollo del conocimiento científico constituye la más fructífera manera de estudiar el conocimiento en general" [Popper (1960), p.188].

epistemológico existiría sólo como consecuencia del racionalismo crítico y este último conllevaría inevitablemente una concepción anarquista del método.

2. Veamos ahora en qué consiste fundamentalmente el racionalismo crítico. Si bien el mismo Popper introduce este término considerándolo un sistema filosófico y un criterio de racionalidad, puede también encontrarse de forma generalizada en otros lugares de la historia del pensamiento, como por ejemplo en los presocráticos[3].

El criterio del racionalismo crítico debe relacionarse tanto con la idea de una demarcación entre ciencia y pseudociencia, como también con la idea de la validez de las teorías científicas. Esto se establece como respuesta ante los criterios verificacionistas, que se basan en la denominada lógica inductiva. [Popper (1963), p. 45].

En *La lógica de la investigación científica* aparece una formulación correspondiente a los inicios del racionalismo crítico[4]. Según ésta, una teoría científica *debe* ser refutable y si un sistema que pretende ser científico no es susceptible de refutación, entonces dicho sistema es pseudocientífico. De la distinción entre ciencia y pseudociencia surge la formulación del criterio de racionalidad desarrollada más tarde en *Conjeturas y refutaciones*[5].

Las teorías científicas deben ser refutables, es decir, debe existir una instancia posible de la experiencia que las contradiga y las refute: un *experimento crucial*.

3. Después de haber revisado las bases del racionalismo crítico, veamos en qué consiste el anarquismo epistemológico. Paul Feyerabend desarrolló el anarquismo epistemológico en respuesta a la gran importancia que estaban recibiendo la razón y los criterios racionales en la filosofía de la ciencia. Si en el racionalismo crítico existen criterios estables según los cuales el conocimiento científico puede progresar, en el anarquismo epistemológico no hay tales criterios de racionalidad, no hay una demarcación lógica que garantice la obtención del conocimiento. Según Feyerabend, los elementos racionalistas no grafican fielmente, por una parte, la historia de la ciencia ni cómo se han establecido las teorías científicas en el pasado. Por otra parte, éstos tampoco son elementos son deseables para la ciencia[6]. Dentro de esta clase de nociones indeseables estarían los criterios del

[3] "La conjetura de que Tales alentaba activamente la crítica en sus discípulos explicaría el hecho de que la actitud crítica hacia la doctrina del maestro se volvió parte de la tradición de la escuela jónica" [Popper (1958), p. 28].

[4] "No exigiré que un sistema científico pueda ser seleccionado […] en un sentido positivo; pero sí que sea susceptible de selección en un sentido negativo por medio de contrastes o pruebas empíricas: ha de ser posible refutar por la experiencia un sistema científico empírico" [Popper (1934), p. 40].

[5] "El criterio de falsabilidad es una solución a este problema de demarcación, ya que dice que los enunciados o sistemas de enunciados, para ser clasificados como científicos, deben ser capaces de estar en conflicto con observaciones posibles o concebibles" [Popper (1963), p. 48].

[6] Refiriéndose a la crítica de Feyerabend en contra del reduccionismo, John Preston afirma que Feyerabend utilliza comúnmente una doble estrategia. "Feyerabend intenta mostrar, primero,

racionalismo crítico [7]. Pero no sólo el criterio de falsación impediría un desarrollo correcto de la ciencia, sino cualquier regla metodológica estricta[8].

Esta afirmación no quiere decir únicamente que muchos criterios metodológicos hayan fallado constantemente en el pasado—lo cual es bastante obvio—sino que toda regla epistemológica es intrínsecamente quebrantable.

4. Veamos ahora cómo es que podemos entender el anarquismo epistemológico como una consecuencia del racionalismo crítico. A pesar de la evidente diferencia que existe entre el racionalismo crítico y el anarquismo epistemológico, existen también algunos aspectos de similitud. Es posible afirmar, incluso, que más allá de las similitudes, el anarquismo epistemológico ha sido *influido* en gran medida por el racionalismo crítico.

4.1) Una de las maneras en que Popper ha influido en Feyerabend está relacionada con la importancia de la distinción entre *falsacionismo ingenuo* y *falsacionismo sofisticado*. Para entender estas nociones es necesario recordar que, según Popper, la refutabilidad de las teorías corresponde a una "refutación de la experiencia" (en el caso de *La lógica de la investigación científica*) y a "eventos concebibles" (en el caso de *Conjeturas y refutaciones*). Esto muestra claramente que las teorías deben ser refutadas por instancias individuales y a la vez experimentales, en el mejor de los casos. El falsacionismo ingenuo posee precisamente la noción de instancia refutadora, que estaría relacionada con un avance lineal de las teorías[9].

Por otra parte, en el falsacionismo sofisticado la refutación no sucede simplemente entre la teoría y un evento singular, sino también entre teorías y series de teorías[10]. Según Lakatos, para el falsacionista ingenuo una teoría científica es falsada si es refutable por un enunciado observacional, mientras que para el falsacionista sofisticado una teoría científica es falsada si entra en conflicto con otra teoría [Lakatos (1978), p. 46]. El problema del progreso científico no corresponde a un proceso lineal, sino a un proceso de series de teorías. Lakatos afirma que existe a veces en Popper una actitud más cercana al falsacionismo sofisticado, pero que el hecho de no distinguir entre teorías y series de teorías le impidió avanzar más en este aspecto [Lakatos (1978), p. 50].

Aunque no se pretende aquí discutir la filosofía de Lakatos, la distinción anterior será de gran importancia para aclarar lo siguiente. En primer lugar, el

---

que la metodología en cuestión no es una descripción exacta de la práctica científica; y luego, que la aplicación de dicha metodología sería indeseable" [Preston (1997), p. 84]. Esta estrategia se encuentra, por ejemplo, en [Feyerabend 1962, p. 67].

[7] "Un principio estricto de falsación […] destruiría por completo la ciencia tal y como la conocemos y nunca le habría permitido empezar" [Feyerabend (1975), p. 162].

[8] "No hay una sola regla, por plausible que sea, y por firmemente basada que esté en la epistemología, que no sea infringida en una ocasión u otra" [Feyerabend (1975), p. 7].

[9] "Los falsacionistas ingenuos sugieren un crecimiento lineal de la ciencia, en el sentido de que las teorías son seguidas de refutaciones poderosas que las eliminan, y tales refutaciones, a su vez, son seguidas por nuevas teorías" [Lakatos (1978), p. 52].

[10] "El falsacionismo sofisticado transforma así el problema de cómo evaluar las teorías en el problema de cómo evaluar las *series de teorías*" [Lakatos (1978), p. 50].

anarquismo epistemológico posee principios similares a la filosofía de Lakatos. Ahora bien, la epistemología de Lakatos está claramente influenciada por el racionalismo crítico y por lo que él llama falsacionismo. Por lo tanto, es posible afirmar que el anarquismo epistemológico también lo está.

La filosofía de Feyerabend posee varios elementos del falsacionismo sofisticado. Un ejemplo de esto es el pluralismo teórico desarrollado por él en su artículo *Explicación, reducción y empirismo* [Feyerabend (1962)], donde afirma que el progreso de la ciencia se produce gracias a la inconsistencia entre las teorías nuevas y las teorías antiguas[11]. Dicho artículo está claramente influenciado por las ideas del racionalismo crítico, lo cual el mismo Feyerabend confiesa.

Así es como las características de la filosofía de Lakatos, relacionadas con los cambios históricos, y con el progreso de la ciencia a partir de series de teorías permiten pensar en las similitudes que hay entre el falsacionismo sofisticado y el anarquismo metodológico. Esta igualdad es algo que admite incluso el propio Feyerabend.[12]

Feyerabend distingue cuatro posturas metodológicas en *La ciencia en una sociedad libre*: a) el racionalismo ingenuo, en el cual está incluído Popper; b) el racionalismo de la dependencia contextual; c) el anarquismo ingenuo; y d) su propio punto de vista. Al explicar este último y la relación que posee su filosofía con las anteriores, Feyerabend se inclina más hacia el racionalismo ingenuo que hacia el racionalismo contextual.

> "Estoy de acuerdo con c*a* [vale decir, la conjunción del anarquismo ingenuo con el racionalismo ingenuo], pero no con c*b* [la conjunción del anarquismo ingenuo con el racionalismo contextual]. Mantengo que todas las reglas tienen sus limitaciones, pero no que debamos proceder sin reglas. Defiendo un enfoque contextual, pero no que las reglas contextuales vayan a *reemplazar* a las reglas absolutas, sino sólo a *complementarlas*. En mi polémica no trato de eliminar las reglas ni tampoco pretendo demostrar su inutilidad. Mi intención es más bien ampliar el repertorio de reglas y sugerir asimismo un nuevo uso de ellas" [Feyerabend (1978), p. 193].

Después de todo, Feyerabend admite estar más de acuerdo con un anarquismo racionalista que con un anarquismo contextualista. La importancia del contexto en el desarrollo de la ciencia no reemplaza la importancia de los criterios de demarcación racionalistas, sino que los complementa y los apoya. Una concepción

[11] "Usuamente, algunos de los principios implicados en la determinación de los significados de los puntos de vista o teorías más antiguas son inconsistentes con las nuevas teorías, que son mejores" [Feyerabend (1962), p. 125].

[12] "Concluyo, pues, que no existe ninguna diferencia susceptible de ser descrita 'racionalmente' entre Lakatos y yo, tomando siempre los criterios de Lakatos como medida de la razón" [Feyerabend (1975), p. 174]. "Lakatos me atribuye a mí un punto de vista psicologista y se atribuye a él mis verdaderos puntos de vista. […] Aparece claro que la creciente separación entre historia de la ciencia, la filosofía de la ciencia y la ciencia misma constituye una desventaja y que debería terminarse con esta separación en interés de las tres disciplinas" [Feyerabend (1975), p. 75].

del conocimiento "contra el método" significa que hay una pluralidad de métodos y reglas que permiten el avance científico, lo que no quiere decir que la carencia completa de todo método sea algo deseable para el conocimiento humano.

4.2) Otra de las influencias del racionalismo crítico sobre el anarquismo epistemológico radica en las concepciones realistas de la ciencia y en la *oposición al instrumentalismo*.

Karl Popper critica enfáticamente el instrumentalismo científico, posición según la cual las teorías científicas no serían más que herramientas predictivas, cuya utilidad sería realizar aportes a las ciencias aplicadas. El racionalismo crítico opta por un *realismo*, es decir, por la idea de que las teorías describen y dicen algo objetivo del mundo[13]. Feyerabend se opone igualmente al instrumentalismo y afirma que el realismo científico es preferible [1964]. Según John Preston, su argumento a favor del realismo es bastante claro[14].

Es cuestionable, en todo caso, si esta defensa del realismo sería válida también con respecto al anarquismo epistemológico. Lo cierto es que antes de sus teorías anarquistas, Feyerabend no era sólo un realista, sino incluso un realista popperiano[15].

4.3) Otra manera importante en que influye el racionalismo crítico al anarquismo epistemológico depende de la *refutabilidad del mismo criterio de racionalidad*. Imre Lakatos afirma que es posible analizar la posibilidad de refutación de los mismos criterios de demarcación mediante la crítica racional, lo cual denomina *metafalsacionismo*[16]. En otras palabras, el metafalsacionismo consiste en aplicar un criterio de racionalidad a otro criterio de racionalidad, idea que Popper nunca desarrolló en un sentido estricto.

Ahora bien, dentro de la filosofía de Feyerabend existen ideas que están muy relacionadas con el metafalsacionismo. Podemos afirmar que el punto de vista del pluralismo metodológico es consecuencia de una crítica racional hacia el racionalismo crítico mismo. Es preciso, sin embargo, no confundir el pluralismo metodológico con el pluralismo teórico. Mientras que el pluralismo teórico proviene del punto de vista contrario al monismo teórico empirista, según el cual una sola teoría no basta para garantizar el conocimiento científico; el pluralismo

---

[13] "No objetamos la afirmación de que todas las teorías científicas son instrumentos, actuales o potenciales. Pero afirmamos que no son *meramente* instrumentos. Pues afirmamos que de la ciencia opdemos aprender algo acerca de nuestro mundo" [Popper (1994), p. 215].

[14] "El progreso científico es deseable; éste es fomentado de mejor forma por una proliferación teórica; y el realismo científico conduce a la proliferación de las teorías, mientras que el positivismo no" [Preston (1997), p. 63].

[15] "No recuerdo haber producido una sola idea que no esté ya contenida en la tradición realista y especialmente en la versión del Profesor Popper" [Feyerabend (1965), p. 251]. En la versión posterior de este artículo de *Realism, Rationalism & Scientific Method*, Feyerabend edita la referencia a Popper.

[16] "En realidad Popper nunca suministró una teoría sobre la crítica racional de las convenciones consistentes. No sólo no responde sino que nunca se plantea la pregunta «¿en qué condiciones abandonaría su criterio de demarcación?»" [Lakatos (1978), p. 161].

metodológico afirma que no existen metodologías únicas que garanticen el progreso científico[17]. Feyerabend aplica constantemente el racionalismo crítico a sí mismo y a los criterios racionales en general con el fin de obtener las consecuencias del pluralismo metodológico[18].

Si el racionalismo crítico es aplicado a sí mismo, en cuanto criterio de racionalidad, entonces se obtiene como consecuencia una inestabilidad de las metodologías de investigación como tales. Aunque parezca extraño esto no implicaría abandonar el racionalismo crítico e incluso lograría asemejarse bastante al pluralismo metodológico.

Para finalizar, quisiera regresar a la pregunta del inicio. *¿Es el anarquismo epistemológico una consecuencia necesaria del racionalismo crítico?* Se ha mostrado que la relación entre uno y otro no es una enorme contrariedad, como se cree por lo general[19]. El anarquismo epistemológico no es una consecuencia necesaria del racionalismo crítico, aunque éste sí es una condición necesaria del anarquismo epistemológico. Existe indudablemente un camino que conduce desde un punto de vista epistemológico hacia el otro. No obstante, y por fortuna, este camino puede corregirse cada vez que la crítica racional lo permita.

Referencias bibliográficas

Feyerabend, P. (1958), "An attempt at a realistic interpretation of experience", en *Realism, Rationalism & Scientific Method*, Cambridge, Cambridge University Press, 1981.

— (1962), "Explicación, reducción y empirismo", en *Límites de la ciencia*, Barcelona, Paidós, 1989.

— (1964), "Realism and instrumentalism: comments on the logic of factual support", en *Realism, Rationalism & Scientific Method*, Cambridge, Cambridge University Press, 1981.

— (1965), "Reply to criticisms", en Cohen, R. S. y Wartofsky, M. (eds.), *Boston Studies in the Philosophy of Science*, New York, Humanities Press, 1965.

[17] John Preston aclara muy bien estas distinciones. Con respecto al pluralismo teórico, afirma que: "[Feyerabend] se opuso constantemente a la exigencia empirista más familiar de aceptar sólo las teorías cercanamente conectadas con la experiencia" [Feyerabend (1997), p. 75]. Este punto de vista estaría ligado a un monismo teórico. Ahora, con respecto al pluralismo metodológico, Preston sostiene que: "El corazón del caso de *Tratado contra el método* es la afirmación de que el monismo metodológico es falso: Feyerabend se ha convertido en un pluralista metodológico" [Preston (1997), p. 170].

[18] "Lo que sostengo es que, dados una regla o un criterio, siempre es posible imaginar un caso que las contravenga. Esto es una 'conjetura audaz', muy querida por los popperianos. Yo jamás he intentado *demostrarla*, pero sí he tratado de hacerla *plausible* aduciendo casos en los que se violan las reglas y los criterios de racionalidad" [Feyerabend (1978), p. 252].

[19] El mismo Paul Feyerabend reconoce que su filosofía es una consecuencia obvia del racionalismo crítico en el siguiente pasaje: "*Todo vale* es una consecuencia práctica obvia de un *racionalismo crítico* como éste, al que Popper solía introducir diciendo que aunque él era un profesor del método científico, no podía actuar de acuerdo con eso, ya que *no existe el método científico*" [Feyerabend (1958), p. 21].

— (1975), *Tratado contra el método*, Madrid, Tecnos, 2000.
— (1978), *La ciencia en una sociedad libre*, México, Siglo XXI, 1998.
Lakatos, I., (1978), *La metodología de los programas de investigación*, Madrid, Alianza, 1989.
Lakatos, I. and Feyerabend, P. (1999), *For and Against Method*, Chicago, Chicago University Press.
Popper, K. (1934), *La lógica de la investigación científica*, Madrid, Tecnos, 1999.
— (1958), "Los comienzos del racionalismo", en Miller, D. (ed.) *Popper: escritos selectos*, México, FCE, 2006.
— (1960), "El desarrollo del conocimiento científico", en Miller, D. (ed.) *Popper: escritos selectos*, México, FCE, 2006.
— (1963), *Conjectures and Refutations*, London, Routledge.
— (1994), *El mito del marco común*, Barcelona, Paidós, 2005.
Preston, J. (1997), *Feyerabend: Philosophy, Science and Society*, Cambridge, Polity Press.
Theocharis T. y Psimopuolos M. (1987), "Where Science has gone Wrong", *Nature*, vol. 329, no. 6140, pp. 595-598.

Por qué el diseño inteligente no puede constituir una teoría científica

Vicente Claramonte
Universitat de València
Vicente.Claramonte@uv.es

Breve reconstrucción del argumento del diseño inteligente

Presenta dos vectores, negativo y positivo.

> (a) Negativo: confirma el diseño inteligente cuanto desautoriza la teoría evolucionista.
>
> (b) Positivo: a partir de la aparente intencionalidad en el ensamblaje de un sistema biótico, puede inferirse que fue inteligentemente diseñado.

El vector negativo (a) adolece de dos inconsistencias.

1ª Falacia *non sequitur*: de las carencias del evolucionismo no necesariamente se sigue que el diseño mejore su fundamento epistemológico. Razonando por analogía, valdría decir que el geocentrismo mejora su fundamento epistemológico como teoría astrofísica porque el heliocentrismo no permite explicar la materia oscura.

2ª Falso dilema: entre teoría evolucionista e inferencia del diseño, como explicaciones únicas y excluyentes sobre el origen y transformación de la vida y las especies, pues existen alternativas, como la teoría transformista de Lamarck sobre la herencia de los caracteres adquiridos y otras. Ni aquéllas son las dos únicas explicaciones, ni tampoco son alternativamente excluyentes, como puedan serlo un enunciado *p* y su negación.

Resultando patente que el vector negativo (a), al argumentar sólo a partir de los defectos de la teoría evolucionista, por sí mismo no aporta fundamento epistemológico alguno a favor del diseño inteligente, centraremos nuestro análisis en el vector positivo (b). Descansa en dos ideas, llamadas complejidad específica y complejidad irreducible. Someteremos la primera a análisis lógico, para comprobar su compatibilidad con el método hipotético-deductivo, y la segunda a análisis empírico, para verificar si existen efectivamente estructuras u organismos irreduciblemente complejos en la naturaleza.

Complejidad específica

Según William Dembski, hay complejidad específica en todo suceso, dotado de un patrón reconocible, con probabilidad de ocurrencia espontánea inferior a 10^{150} [CE = $p < 10^{150}$], bajo cuyo umbral puede asumirse que no fue producido por azar ni causa natural, sino diseñado o producido por causalidad inteligente (Dembski, 1999: p. 47 y ss.).

Su complemento teórico es el llamado filtro explicativo, según el cual, un suceso sólo puede ocurrir por tres tipos de causas:

1ª	Necesidad	= probabilidad alta o regularidad
2ª	Azar	= probabilidad intermedia
3ª	Diseño	= $p < 10^{150}$

Si pueden descartarse las dos primeras, debe admitirse la causalidad por diseño.

2.1) Objeciones: al filtro explicativo

1ª Prescinde del método científico. No procede a una prueba por casos completa de las hipótesis alternativas posibles, sometiéndolas todas a idéntico análisis. Argumenta contra dos de las hipótesis definidas y, tras descartarlas, concluye aceptando la tercera sin someterla al mismo análisis. Estamos ante un método probatorio de descarte por casos, pero aplicado de modo incompleto.

2ª Selección incompleta de hipótesis. Aplicado a nivel microbiológico para descartar la necesidad[1] —en teoría evolucionista representada por la selección natural—, omite una hipótesis crucial: la posible combinación entre alternativas. Pues la combinación de azar (variabilidad) y necesidad (selección natural), sí genera información genética nueva, al activarse una mutación genética ($azar_1$) que ofrece una ventaja adaptativa en un entorno cambiante ($azar_2$), y luego los individuos beneficiados y su descendencia son favorecidos por selección natural (necesidad).

2.2) Objeciones: a la complejidad específica

3ª Tautología sin valor informativo real sobre los fenómenos naturales. No ocurre *ex natura*, sino *more definitio*; no aparece en la naturaleza, sino que es un postulado introducido sin demostración alguna por la definición dembskiana. Por ello, el valor veritativo de los enunciados construidos sobre complejidad específica es nulo en cuanto a información empírica aportada acerca del mundo natural.

4ª Subsidiariedad. Para bloquear la causalidad evolucionista, basada en la combinación de mutación aleatoria y selección natural antes descrita, el concepto de complejidad específica termina remitiendo al de complejidad irreducible, y por tanto su fuerza argumentativa es subsidiaria de éste.

Resultando patente la inconsistencia de la complejidad específica, el fundamento epistemológico del diseño inteligente como teoría científica en Biología depende exclusivamente de la complejidad irreducible.

Complejidad irreducible

Esta idea es un contra-concepto deducido por Michael Behe a partir del siguiente párrafo, «*Si pudiera demostrarse que ha existido un órgano complejo que no pudo*

[1] Para Dembski, los organismos son complejamente específicos en el ámbito microbiológico, es decir, están causados por diseño inteligente —única alternativa capaz de generar información genética nueva— y no por azar o necesidad.

haber sido formado por numerosas y ligeras modificaciones sucesivas, mi teoría fracasaría por completo», escrito por Charles Darwin en *El origen de las especies* (Darwin, 1970; capítulo IV, "Dificultades de la teoría"). Behe afirma aludiendo a dicha cita:

> «Un sistema que cumple el criterio de Darwin es aquél que exhiba una complejidad irreducible. Por complejidad irreducible entiendo un sistema simple compuesto de varias partes que interactúan contribuyendo a una función básica, y donde la supresión de cualquiera de estas partes hace que el sistema cese de funcionar con eficacia» (Behe 2000; "Máquinas moleculares: apoyo experimental para la inferencia del diseño").

Propone después tres supuestos de complejidad irreducible en la naturaleza: flagelo bacteriano, cascada coaguladora sanguínea y sistema inmunitario. Dada su estructura tan compleja y sus elementos integrantes tan funcionalmente insustituibles, afirma que estos sistemas no pudieron formarse por leves modificaciones sucesivas, como predice la teoría evolucionista, y por tanto carecen de precedente evolutivo y fueron creados abruptamente.

En consecuencia, la confirmación o refutación de la inferencia del diseño puede comprobarse verificando si los elementos de esa presunta complejidad irreducible son o no sustituibles sin un colapso global de la funcionalidad del sistema complejo, y por ello, si pudieron o no formarse gradualmente por acumulación de ligeras y sucesivas modificaciones. Todo lo cual, se concreta en la existencia o inexistencia en la naturaleza de precedentes evolutivos de los sistemas propuestos por Behe como ejemplos de complejidad irreducible.

Pero antes de verificar la existencia de órganos o estructuras irreduciblemente complejas en el sentido definido por Michael Behe, establezcamos un breve paréntesis en esta discusión para introducir ciertas reflexiones sobre Filosofía del Lenguaje, coadyuvantes a valorar el estatus epistemológico de la inferencia del diseño como teoría científica, e inmediatamente iniciaremos el rastreo empírico de los sistemas bióticos que permitan confirmar o descartar la existencia de complejidad irreducible en la naturaleza.

3.1) Conceptos vacíos y actos epistémicos fallidos

Todo concepto es el producto de una doble operación mental, consistente en la abstracción de las propiedades empíricas que diferencian a un conjunto de objetos, y a la vez, en la síntesis de las propiedades inteligibles que los asemejan, y cuyo resultado final conlleva subsumir intelectualmente un objeto físico o abstracto mediante un término dotado de expresión fonética y gráfica. Dos de las propiedades básicas de un concepto son el significado —contenido semántico—, y la extensión, —conjunto de entes subsumidos por su significado. Por ejemplo, en el concepto "*día de la semana*", su significado sería "*período natural y consecutivo de 24 horas agrupado en grupos de siete elementos*", y su extensión, el conjunto integrado por los elementos "*lunes, martes, miércoles, jueves, viernes, sábado y domingo*".

Todo objeto puede subsumirse bajo un concepto. Este enunciado es verdadero incluso en el supuesto diametralmente adverso de entes sustraídos a las

capacidades cognoscitivas; siempre podrían subsumirse bajo el concepto "*entidades incognoscibles*", y después definirse su extensión E, es decir, el conjunto de entidades incognoscibles abarcadas por su significado: E (eeii) = $\{ei_1, ei_2, ei_3 \ldots ei_n\}$. En cambio, no todo concepto subsume a un ente físico. Por ejemplo, el significado del concepto "*habitante solar*", es perfectamente comprensible, "*ser cuya morada es el Sol*", pero al carecer de referente empírico, es imposible vincularlo a un ente físico, por lo cual constituye una clase sin elementos y su extensión sólo puede ser un conjunto vacío, E (hs) = $\{\emptyset\}$.

Por ello, llamamos vacíos a estos conceptos que no subsumen ente físico alguno, carecen de referente empírico y su extensión es un conjunto vacío. La Historia de la Ciencia relata una larga tradición de problemas epistemológicos derivados del uso espurio de conceptos vacíos. Baste recordar los perdurables e inútiles programas de investigación destinados a localizar éter en Física, flogisto en Química o calórico en Termodinámica. La operación intelectual resultante de una conceptuación vacía puede considerarse un acto epistémico fallido (Díez y Moulines 1997: p. 93), pues al carecer de referente empírico, el concepto resultante no subsume ente físico alguno, y por tanto es incapaz de producir conocimiento auténtico sobre la naturaleza. En consecuencia, el empleo de conceptos vacíos constituye un acto epistémico fallido para generar conocimiento empírico, y por tanto, es inhábil para articular una teoría sobre el área cognitiva característica de las ciencias naturales.

Sentado lo anterior, volvamos de nuevo a la inferencia del diseño inteligente, para comprobar si hallamos en el flagelo de las bacterias, la cascada de coagulación sanguínea y el sistema inmunitario, estructuras u órganos constituyentes de un referente empírico que pueda subsumirse bajo la idea de complejidad irreducible definida por Michael Behe.

3.2) Rastreo empírico de la presunta complejidad irreducible

1º Flagelo bacteriano.[2] La literatura experta demuestra múltiples homologías entre éste y una versión suya simplificada, una subestructura del flagelo que mantiene (otra) funcionalidad pese a perder algún componente, por lo cual se admite una relación evolutiva entre ambos. Entre 1990 y 2003, diversos estudios identificaron un subsistema de la estructura bioquímica del flagelo enteramente operativo, llamado sistema secretor Tipo III o inyectosoma. Aunque se investiga aún si el flagelo bacteriano evolucionó en el sistema secretor Tipo III, éste en aquél o ambos a partir de un ancestro común, su relación evolutiva se considera evidencia científica demostrada.

2º Cascada coaguladora.[3] Estudios comparativos muestran que la sangre de delfines y ballenas coagula pese a perder parte de la cascada de coagulación, según demostraron pruebas moleculares en 1998. En especies como el pez globo, la

[2] Véase por ejemplo Michiels (1990), Braun (2001), Burr *et al.* (2002), Gavín (2003), Matzke (2003); etc.

[3] Véase por ejemplo Xu y Doolittle (1990), Miller (1999), Doolittle (2001), Davidson *et al.* (2003), Forrest y Gross (2007); etc.

sangre coagula pese a que la estructura bioquímica de la cascada coaguladora aparece incompleta; ergo, no existe complejidad irreducible. Además, ciertas especies actuales de pepinos marinos cuentan con un gen productor de una proteína similar al fibrinógeno —responsable último de la coagulación—, pero sin función metabólica coaguladora. Esto indica que la modificación evolutiva de una proteína preexistente en invertebrados ancestrales, con función distinta, constituye el origen químico de las moléculas de fibrinógeno que hoy implementan la coagulación en los vertebrados. Idéntico contexto hematológico aparece en otras especies, como las lampreas y el pez locomotora.

3º Sistema inmunitario.[4] La presencia de rasgos evolutivos en la conformación del sistema inmunitario puede rastrearse en sus dos tipos, innato y adaptativo.

(a) Innato: está integrado por la cascada proteínica complementaria y ésta desempeña además un papel clave en el adaptativo, mediante la modulación y modificación de la respuesta de las células T. Por ello, la estructura constituyente de complejidad señalada por Behe en el sistema inmunitario sí es reducible a otra subestructura, llamada cascada proteínica complementaria, que resulta funcional pese a reducirse el número de sus componentes.

(b) Adaptativo: aunque probablemente fue desarrollado por los primeros vertebrados, muchas especies actuales utilizan mecanismos precursores de funciones desempeñadas por el sistema inmunitario específico de los vertebrados mandibulados. Tales mecanismos precursores incluso aparecen en formas de vida mucho más simples, como las bacterias, las cuales emplean un único sistema defensivo, llamado de restricción y modificación, para afrontar patógenos víricos o bacteriófagos.

Ineptitud del diseño inteligente como teoría en una ciencia empírica

Recapitulando los apartados anteriores, podríamos extraer dos conclusiones:

1ª) Según consta en la literatura científica especializada, puede afirmarse la existencia de subestructuras bioquímicas que muestran precedencia evolutiva en los ejemplos de presunta complejidad irreducible propuestos por Behe como apoyo empírico para la inferencia del diseño inteligente; flagelo bacteriano, coagulación sanguínea y sistema inmunitario. Por tanto, como predice la teoría evolucionista, estamos ante estructuras u órganos formados durante la historia evolutiva de la biótica con numerosas y ligeras modificaciones sucesivas.

2ª) La idea de complejidad irreducible carece de referente empírico, su conceptuación no subsume ente físico alguno y genera un concepto vacío. Cuando conceptos vacíos como "éter", "flogisto", "calórico" o "complejidad irreducible" intentan aplicarse a objetos, constituyen un acto epistémico fallido. Propenden a una filosofía verbalista que no fundamenta el razonamiento en los conceptos sino en las palabras, y al no subsumir ningún ente localizable en una región

[4] Véase por ejemplo Bickle y Krüger (1993), Beck y Habicht (1996), Eason *et al.* (2004), Litman *et al.* (2005), Rus *et al.* (2005), Cooper y Adler, (2006), Pancer y Cooper (2006); etc.

espaciotemporal, son inútiles para articular una teoría en el ámbito de una ciencia empírica como la Biología.

Trasladado nuestro análisis a un esquema premisas-conclusiones, obtendríamos:

P_1 En ciencias empíricas, es imposible elaborar una teoría mediante conceptos vacíos.

P_2 El diseño inteligente se construye a partir del concepto "complejidad irreducible".

P_3 "Complejidad irreducible" es un concepto vacío, como muestra la literatura científica.

C_L El diseño inteligente no permite construir una teoría científica en Biología.

Un discurso propuesto como científico sin serlo es sólo pseudociencia.

Referencias bibliográficas

Beck, G. y Habicht, G. (1996), "Immunity and the Invertebrates", *Scientific American Magazine*, noviembre, pp. 60-6. New York.

Behe, M. (2000), *La caja negra de Darwin. El reto de la Bioquímica a la evolución*, Barcelona, Andrés Bello.

Bickle, T. y Krüger, D. (1993), "Biology of DNA restriction", *Microbiological Reviews*, 57 (2): 434-50.

Braun, P. *et al.* (2001), "Maturation of *Pseudomonas aeruginosa* Elastase. Formation of the Disulfide Bonds". *The Journal of Biological Chemistry*, vol. 276 (28): 26030-5.

Burr, S. *et al.* (2002), "Evidence for a Type III Secretion System in *Aeromonas salmonicida* subsp. *salmonicida*". *Journal of Bacteriology*, 184 (21): 5966-70.

Cooper, M. y Alder, M. (2006), "The evolution of adaptive immune systems". *Cell*, 124 (4): 815-22.

Darwin, CH. (1970), *El origen de las especies*. Barcelona, Zeus.

Davidson, C. *et al.* (2003), "Molecular evolution of the vertebrate blood coagulation network." *Thrombosis and Haemostasis* 89 (3): 420-428.

Dembski, W. (1999), *Intelligent Design: The Bridge Between Science & Theology*. Westmont, InterVarsity Press.

Diez, J. y Moulines, C. (1997), *Fundamentos de Filosofía de la Ciencia*. Barcelona, Ariel.

Doolittle, R. (2001), "Crystal Structure Studies on Fibrinogen and Fibrin", *Annals of the New York Academy of Sciences*, 936 (1): 31-43.

Eason, D. *et al.* (2004), "Mechanisms of antigen receptor evolution", *Seminars in Immunology*, 16 (4): 215-26.

Forrest, B. y Gross, P. (2007), "Biochemistry by design". *Trends in Biochemical Sciences*, 32 (7).

Gavín, R. (2003), *Caracterización genética y fenotípica del flagelo de Aeromonas*, Tesis doctoral: Universidad de Barcelona, Departamento de Microbiología.

Litman, G. *et al.* (2005), "Reconstructing immune phylogeny: new perspectives." *Nature Reviews of Immunology*, 5 (11): 866-79.

Matzke, N. (2003), "Evolution in (Brownian) space: a model for the origin of the bacterial flagellum", <http://www.talkdesign.org.faqs/flagellum.html>.

Michiels, T. *et al.* (1990), 'Secretion of Yop proteins by *Yersiniae'*. *Infection and Inmunity*, 58 (9): 2840-9.

Miller, K. (1999), *Finding Darwin's God: A Scientist's Search for Common Ground Between God and Evolution*, New York, Harper Collins Publishers.

Pancer, Z., y Cooper, M. (2006), "The evolution of adaptive immunity." *Annual Review of Immunology*, 24: 497-518.

Rus, H. et al. (2005), "The role of the complement system in innate immunity", *Immunologic Research*, 33 (2): 103-12.

Xu, X. y Doolittle, R. (1990), "Presence of a vertebrate fibrinogen-like sequence in an echinoderm", *Proceedings of the National Academy of Sciences*, 87: 2097-2101.

Holismo semántico y leyes constitutivas*

José Luis Falguera
Universidad de Santiago de Compostela
joseluis.falguera@usc.es

I. Quine (1951) adoptó la tesis del holismo epistémico. Una formulación del mismo ampliamente aceptada reza así:

> *(I)* No es posible la contrastación –confirmación o desconfirmación– de una hipótesis científica aislada; una contrastación presupone un haz de hipótesis, cualquiera de las cuales puede ser considerada inadecuada en caso de desconfirmación.

De lo que resulta una consecuencia metodológica:

> *(II)* Partiendo de una hipótesis –considerémosla potencialmente inadecuada–, dada una desconfirmación, podemos mantener dicha hipótesis si introducimos cambios en cualquier(cualesquiera) otra(s) hipótesis del haz del que forma parte aquélla, a fin de intentar superar la desconfirmación.

II. Quine dio un paso más, estableciendo su tesis del holismo semántico. El punto de encuentro entre holismo epistémico y holismo semántico para Quine (1969) se apoya en un criterio verificacionista del significado.

La propuesta de holismo semántico quineana de las expresiones teóricas (en su versión más radical en lo que concierne a unidad de significación) consiste en un argumento como el siguiente:

1. Criterio verificacionista: Tiene significación (empírica) aquella unidad lingüística que es susceptible de contrastación (es decir, cuyo sostenimiento produce una diferencia en la experiencia posible).
2. Holismo epistémico: Cada enunciado teórico aislado no es susceptible de contrastación (confirmación o desconfirmación), es el conjunto de enunciados teóricos que conforman la totalidad de la ciencia lo que es susceptible de contrastación.
3. Por lo tanto (de 1 y 2), holismo semántico (*á la Quine*): los enunciados teóricos tomados individualmente uno por uno no son unidades de significación (empírica), lo es (cada formulación lingüística alternativa de) la totalidad de la ciencia.

* Este trabajo participa en los proyectos de investigación del Ministerio de Ciencia e Innovación (España) HUM2006-04955/FISO, y de la Agencia Nacional de Promoción Científica y Tecnológica (Argentina) PICT Redes 2006 Nº 2007.

Frente a los enunciados teóricos, cada enunciado observacional sí tiene significación (empírica) por sí mismo.

III. La propuesta quineana de holismo semántico encierra una suerte de incoherencia interna. El argumento está extraído de Fodor y Lepore (1992, cap. 2), aunque formulado de manera diferente y con otras pretensiones. Aquí con el mismo se pretende cuestionar el holismo semántico *á la Quine* (aunque no toda propuesta de holismo semántico para expresiones científicas). Veámoslo:

- Cualquier cambio en los enunciados teóricos de C (una formulación de la ciencia) para obtener C* (una formulación alternativa de la ciencia que modifica a C) que conlleve cambios en los enunciados observacionales que se siguen como consecuencias, supone un cambio de significado (C y C* tienen significados diferentes).
- Dado que no se desconfirma un determinado enunciado teórico α de C, sino todo C –siendo α la hipótesis potencialmente inadecuada–, α podría *mantenerse* si se establecen modificaciones de otros enunciados teóricos de C para obtener C* y evitar la desconfirmación [consecuencia metodológica *(II)* del holismo epistémico].
- Pero, al sostener la formulación quineana del holismo semántico, ¿qué cabe entender por *mantener* un enunciado teórico α, cuando con el cambio de algún(os) otro(s) enunciado(s) teórico(s) para obtener C* cambian los enunciados observacionales que se siguen como consecuencias de C*, es decir, cambia el contenido empírico de C* con respecto al que tenía C?
- Mantener la mera expresión sintáctica no es garantía de que se ha mantenido lo que se precisaba, ya que el cambio de enunciados observacionales que se siguen como consecuencias de C* supone el cambio de significado de C* con respecto a C. Por tanto, nada garantiza que la expresión sintáctica mantenida (como componente de C*) siga estableciendo lo mismo (signifique lo mismo) que lo que establecía (como componente de C). Y si no significa lo mismo, ¿qué relevancia tiene decir que mantenemos una expresión sintáctica cuando trata de algo completamente diferente?
- Obviamente, desde el punto de vista epistémico sí que es relevante que se mantenga algo más que la mera expresión sintáctica si queremos decir que se mantiene la misma hipótesis.
- Si aceptamos que mantener un enunciado teórico α, en caso de desconfirmación, es mantener la expresión sintáctica con un contenido intensional que perdura con el cambio de C a C*, nos encontramos con que aceptar tal cosa es tanto como poner en cuestión la versión quineana del holismo semántico para expresiones científicas, no sólo por la mera aversión quineana a las entidades intensionales, sino porque habría que rechazar su modalidad de criterio verificacionista de significación.

IV. Kuhn, especialmente en sus últimos trabajos (p.e., 1989; 1993), defiende un holismo semántico para las expresiones científicas diferente del de Quine, pese a que las referencias que el primero hizo de las propuestas semánticas del segundo puedan dar lugar a pensar lo contrario. Hay aspectos que ya muestran una concepción semántica diferente entre ambos. Así Kuhn:

- rechaza la invarianza de significado de los enunciados observacionales, y con ello que puedan servir como elemento de contrastación neutral para cualesquiera pares de teorías;
- no asume, pues, el criterio verificacionista de significado;
- no rechaza las entidades intensionales.

Pero, además, Kuhn plantea un holismo semántico local. Éste sólo establece la dependencia de *algunos de los términos característicos* de una teoría con respecto a la misma. En concreto:

- Dada una teoría T, Kuhn distingue entre vocabulario específico de T y vocabulario disponible previamente a T (vocab. específico de T + vocab. disponible previamente a T = vocab. característico de T).
- Los significados de los términos previamente disponibles a T se alcanzan con independencia de T.
- Los significados de los términos específicos de T dependen de T.
- La dependencia de los términos específicos de una teoría T respecto a la misma supone interdependencia de los significados de esos términos en el marco de T.
- Kuhn plantea que la interdependencia semántica de los términos específicos de una teoría T se asegura gracias a generalizaciones con un rol estipulativo o constitutivo (en cuanto se consideran en relación a algunas aplicaciones pretendidas).
- Sin embargo, también plantea que el significado de cada uno de los términos específicos de T se adquiere/aprende con generalizaciones con rol estipulativo y con algunas otras con rol empírico (en cuanto se consideran en relación a algunas aplicaciones pretendidas).
- Las generalizaciones con rol estipulativo son finalmente presentadas como 'sintético *a priori* no-absolutas'.

Así pues, el holismo semántico local que propugna Kuhn tiene un sesgo racionalista en su formulación, frente a la orientación empirista del de Quine. En todo caso las consideraciones previas abren dos problemas: (i) el de la elucidación de las generalizaciones sintético *a priori* no absolutas de una teoría; (ii) el de cuán acotado es el holismo semántico local, es decir, si la dependencia semántica de los términos específicos (su significado) se debe acotar a las generalizaciones sintético *a priori* no-absolutas o si por el contrario debe contemplar generalizaciones con rol empírico (a la luz de lo que Kuhn señala acerca de cómo se adquieren/aprenden los términos específicos de una teoría).

V. Respecto a la elucidación de las generalizaciones sintético *a priori* no-absolutas propongo que corresponden a lo que la metateoría estructuralista denomina leyes fundamentales. Una teoría compleja, conforme a esa corriente, se caracteriza por disponer de una ley fundamental y varias leyes especiales. Las leyes de una determinada teoría estarían organizadas jerárquicamente conformando una imagen arbórea (como en la Figura1), de manera tal que, si cada nudo del árbol de leyes de una teoría corresponde a una ley, habría un nudo común a todas las ramas (l_f), que representa a la ley fundamental de la teoría T (LF-T), y cada rama desembocaría en un nuevo nudo ($l_{i.j}$ ó $l_{i.j.k}$ ó …) que representa a una ley especial de la teoría T, la cual restringe (especializa) a las que le preceden hasta el nudo común. Una ley especial presupone las que le preceden en el árbol (aquellas otras de las que es una especialización). De esa imagen es relevante que las leyes más especializadas se consideran (normalmente) en relación con un menor conjunto de aplicaciones que las menos especializadas (aquellas restringen más que éstas); y sólo la LF-T se considera en relación con todas las aplicaciones pretendidas de la teoría T.

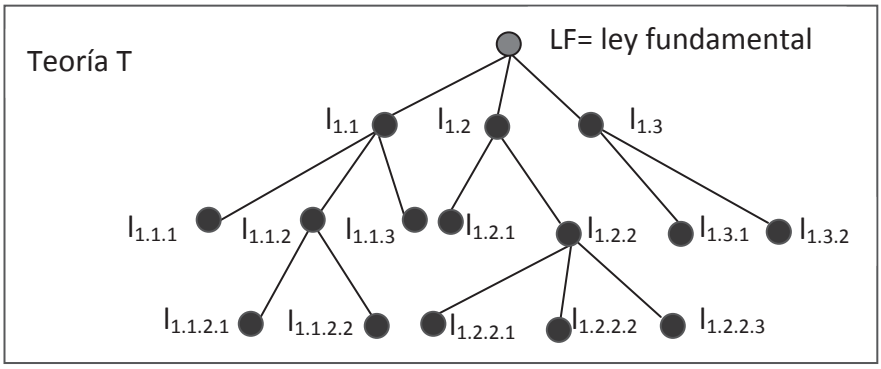


Figura 1

La literatura estructuralista indica no disponer de un criterio con condiciones suficientes y necesarias de lo que es una LF-T, pero al menos recoge algunos rasgos que parecen ser condiciones necesarias:

- El carácter (casi)sinóptico: la LF-T incluye (casi) todos los conceptos característicos de T, conectándolos de manera esencial.
- El carácter cuasi-vacuo: que puede explicarse, según Moulines (1978), por la forma lógica (cuantificación existencial sobre variables de orden superior o funcionales). Dicha forma lógica supone que una LF-T es altamente irrestricta e irrefutable. Con ello se captaría el rasgo de *a priori* no-absoluto que cabe atribuirle.

No obstante, aceptando los rasgos mencionados, tal vez pueda considerar que una formulación invertida del criterio de teoricidad de esta corriente proporciona un criterio de ley fundamental. Éste diría que: la ley fundamental de T es aquella que está presupuesta en toda determinación de al menos un término de T.

VI. Respecto al problema de cuán acotado es el holismo semántico local encontramos que Gähde (1990), desde la metateoría estructuralista, ha dado cuenta de dos condiciones que afectan a los términos T-teóricos de una teoría T que parecen avalar un holismo semántico poco acotado. (Para lo que sigue equiparemos términos específicos de T a términos T-teóricos, y términos de T previamente disponibles a términos T-no-teóricos.) Las dos condiciones son:

> (1) La mera LF-T no determina unívocamente valores para sus términos T-teóricos, una vez dados los valores para los términos T-no-teóricos de las aplicaciones pretendidas.
>
> (2) Un término T-teórico sólo puede ser unívocamente determinado, una vez dados los valores para los términos T-no-teóricos de ciertas aplicaciones pretendidas, gracias a la LF-T y alguna(s) ley(es) especial(es) apropiada(s).

Ambas condiciones podrían interpretarse como una corroboración de que el significado de un término T-teórico no depende sólo de la LF-T, sino también (al menos) de leyes especiales. El argumento establecería que:

- La determinación unívoca de un término T-teórico requiere no sólo de la LF-T sino también de leyes especiales.
- Si la determinación unívoca de un término T-teórico requiere no sólo de la LF-T sino también de leyes especiales, entonces la extensión efectiva total de ese término no la determina sólo la LF-T sino también las leyes especiales.
- El significado de un término científico se suele concebir como una regla para determinar su extensión efectiva.
- Luego, el significado de un término T-teórico no depende sólo de la LF-T sino también de las leyes especiales.

Pero el holismo semántico de las expresiones científicas poco acotado encierra una suerte de incoherencia análoga a la del holismo semántico *á la Quine*. Veámoslo:

- Una teoría cambia con el tiempo, porque cambian las aplicaciones pretendidas y también porque cambian algunas leyes especiales.
- Si el significado de un término T-teórico depende de todas las leyes de T, el significado de ese término cambia con el desarrollo de T.
- Pero si el significado de un término T-teórico cambia con el desarrollo de T, lo que se mantiene con el desarrollo de T es la mera expresión sintáctica.
- Parece que no tiene mucho sentido hablar de la identidad de la teoría cuando cambia el significado de sus términos teóricos.
- Luego, el significado de un término T-teórico no puede depender de todas leyes de T.

Asumo el holismo epistémico que propugna la metateoría estructuralista, y con él que la extensión efectiva total de un término T-teórico depende de todas las leyes de T. Lo que cabe revisar es la consideración de que el significado de un término

científico sea una regla para determinar su extensión efectiva total. Tendríamos que el cambio de extensión efectiva total de un término científico *no* necesariamente supone su cambio de significado. Así, debemos aceptar un holismo semántico de las expresiones científicas acotado, por el que el significado de los términos T-teóricos sólo dependen de su ley fundamental.

Referencias bibliográficas

Fodor, J. A. y Lepore, E. (1992), *Holism. A Shopper's Guide,* Oxford, Blackwell.

Kuhn, T. S. (1989), 'Possible Worlds in History of Science', en Allén, S. (ed.) *Possible World in Humanities, Arts and Sciences*, Berlin, De Gruyter, pp. 9-32.

— (1993), 'Afterwords', en Horwich, *World Changes. Thomas Kuhn and the Nature of Science,* Cambridge (Massachusetts), The MIT Press, pp. 311-341.

Gähde, U. (1990), 'On Intertheoretical Conditions for Theoretical Terms', *Erkenntnis*, 32: 215-233.

Moulines, C. U. (1978), 'Cuantificadores Existenciales y Principios-Guía en las Teorías Físicas', *Crítica* 10, pp. 59-88.

Quine, W. V. (1951), 'Two Dogmas of Empiricism', en *Philosophical Review*, 60: 20-43. (Vers. util.: 'Dos Dogmas del Empirismo', en Quine (1953) *From a Logical Point of View.* Cambridge, Harvard Univ. Press, pp. 49-81).

— (1969), 'Epistemology Naturalized', en Quine (1969), *Ontological Relativity and Other Essays.* New York, Columbia Univ. Press, pp. 69-90.

Clinical and experimental practice in psychology: kinds of inferences

Nicolò Gaj and Giuseppe Lo Dico
Catholic University of Milan
nicolo.gaj@unicatt.it / giuseppe.lodico@unicatt.it

Recently, psychologists' attention has been drawn to the procedures of theory construction, that is to say, to the processes involved in the quest for sound explanations of psychological phenomena [see, for example, Haig (2005); Johnson-Laird (2006); Capaldi & Proctor (2008); *Journal of Clinical Psychology (JCP)* – Special Issue (2008)]. It is a matter of fact that the various areas of psychology have their own peculiarities but nonetheless they also show some commonalities regarding the study of the phenomena they deal with. In particular, clinical and experimental practices have different methods, objects of inquiry and areas of application but they seem to share some inferential procedures in order to improve their results. In the Special Issue of the *JCP* some authors debate upon the reliability of one of them: the abductive method. Abduction is commonly defined as a kind of inference going from data to a hypothesis that best accounts for them [Josephson & Josephson (1994)]. Beginning from this definition, the psychologist Haig develops an abductive theory of scientific method (ATOM) that comprises a set of specific strategies oriented to detect empirical phenomena and construct explanatory theories [Haig (2005)]. It has three methodological phases:

> (1) Theory generation by means of exploratory factor analysis (EFA). EFA is a fundamental statistical tool in those areas of psychology in which underlying components (such as personality traits or other psychological constructs) are suspected to be involved but difficult to discern [Reber & Reber (2001)].
>
> (2) Theory development through analogical modelling (AM). AM is a theory-construction strategy used by the scientist once the hidden factors have been identified through EFA. By using this theoretical strategy, the scientist builds a model by describing these factors and their interrelations in terms of what is already familiar and well understood [Haig (2005)].
>
> (3) Theory appraisal by using the criteria of explanatory coherence (EC) proposed by the philosopher Thagard (1978, 1988, 1992, 2000). EC is an index that evaluates the internal features of a theory and that allows the scientist to choose the best explanation among the different ones at disposal. It consists of three internal criteria: explanatory breadth (a theory is more explanatorily coherent than its rivals if it explains more phenomena), simplicity (a theory is preferable to its rivals if it makes fewer special assumptions), and analogy (a theory is more coherent if it is

supported by analogy to theories that were already found credible) [Haig (2008)].

According to ATOM, (1) chronologically and conceptually precedes (2). Instead, (3) is transversal to all the phases. In fact, theory appraisal begins at (1), continues at (2), and is conclusively carried out in (3). Further, every phase is considered to be abductive in nature, even if the character of the abductive inference is different for each case [Haig (2008)]. In fact, in (1) the existence of new theoretical entities is inferred, in (2) a model of the properties of these theoretical entities is built through analogy, and in (3) a systematic evaluation of competing theories is made.

There are two controversial issues regarding abduction as a scientific method of psychological inquiry. The first concerns the possible weaknesses of ATOM that can expose it at risk of losing its normative status [Romeijn (2008)]. The philosopher Romeijn maintains that this risk is due to one of the most welcomed feature of ATOM: an open eye on what psychologists really do. Although appreciated, this feature is argued to be responsible for the inadequacy of ATOM in providing a sound recipe for properly justifying facts. ATOM would describe what happens in professional research practice without being a guide to the scientific enterprise. In other words, ATOM as it stands fails to answer to a critical question in philosophy of science: how can we demarcate a scientific method of justification from non-scientific ones?

The second problem deals with the fact that the contributors of the Special Issue debate seem to share a negative opinion about the validity of the H-D method, which has to be substituted by any version of the abductive one. According to the H-D method, a certain number of empirical predictions should be made from a single theory. If such predictions match what is observed, then those observations confirm the theory [Bird (1998)]. However, it is important to note that, as well as for the missed distinction between normative and descriptive considered above, none of the contributors make any distinction between the context of discovery (CoD) and the context of justification (CoJ). Such a lack leads to consider the H-D method to be incompatible with abduction. In fact, although Haig seems to implicitly argue that ATOM is a more reliable alternative to the traditional scientific methodology [Haig (2005, 2008b)], it is not clear whether it only deals with the formalization of the discovery enterprise or if it even deals with the justification of hypotheses. In other words, the traditional distinction between the two contexts is replaced by a general notion of theory construction in abductive terms. Thus, if the distinction between CoD and CoJ is not assumed, it is clear that the explanatory role of H-D method is necessarily ignored or declared to be weak [Capaldi & Proctor (2008)].

These two problems are clearly evident in a recent article of the psychologists Capaldi and Proctor (2008). In this contribution, they argue that the main flaw of the H-D method is the fact that it permits to test only a single theory in isolation and not in comparison with other competing ones. According to them, this leads to three weaknesses (that should be solved by the abductive one):

(a) H-D method does not work in the early stages of theory development.

(b) H-D method is open to the formal fallacy known as 'affirming the consequent' (the problem of under-determination of theory by data).

(c) H-D method can easily lead to commit a Type I Error (α), that is to say, to reject a theory when it is correct. This is because, if the experimental results fail to confirm the theory, it may not necessarily indicate that the theory is not correct.

As we can see, the first two points seem to suffer of the missed distinction between CoD and CoJ, whereas the third of that between normative and descriptive. About (a), it is true that any version of the H-D method does not work in the early stages of theory development, contrary to abduction. But this can be considered a fallacy of the former method and a virtue of the latter only if the previous distinction is not assumed. As we pointed out above, this distinction is crucial at the normative level for defining the specific roles of the different methods of investigation. About (b), it is true that the H-D method does not work in cases of under-determination, contrary to abduction. But this can be considered a fallacy of the former over the latter only if it is not defined the specific role of the methods of inquiry. Instead, point (c) deals with the application of the H-D method in the scientific practice. Although at the normative level this method prescribes to refuse a theory when a single prediction does not match with the results, at the practical or descriptive level the scientist does not tend to follow this prescription. Many disconfirmations (maybe those coming from independent scientists and laboratories) can be a very important index at least for putting the theory in question.

In conclusion, we argue that abduction is incompatible with the H-D method only if we do not consider the distinction between the CoD and CoJ, as most of the contributors to the Special Issue debate do. In ATOM alternative hypotheses are generated (trough EFA), developed (trough AM) and appraised (trough EC criteria) in order to attain to the best explanation of the data. In contrast, the crucial point of the H-D method is the following: if the predictions of the theory considered match what is observed, then those observations give a confirmation to the theory. It seems clear to us that ATOM can be considered as a useful conceptualization of what happens in CoD. The upshot of ATOM, a hypothesis generated and assessed through the three phases in CoD, is then compared in CoJ with data, specifying the conditions under which it (premise) deductively implies the data (conclusion).

On the basis of the distinction between CoD and CoJ, we propose an integration between a version of ATOM [Haig (2005); Vertue & Haig (2008)] and the H-D method.

1) Individuation of the object of inquiry (IO).

2) Novel hypotheses generation (NHG). This process can be statistically worked out by EFA, generating competing hypotheses about the object.

3) Inference to the best explanation (IBE). The upshots of NHG are evaluated in terms of EC.

4) Empirical justification (EJ) through H-D method. The best hypotheses selected at stage (3) are singularly compared with data. Their validity is ascertained assessing how they fit with empirical data.

(1), (2) and (3) can be considered as different processes that constitutes CoD. Hypotheses are generated and conceptually assessed in terms of internal criteria. Then, in (4), hypotheses are compared with the world: this process constitutes what traditionally has been called CoJ[1].

The model proposed permits to face a controversial issue in psychological literature: the fact that experimental practice is often seen to be different from the clinical one. But why these two contexts are commonly considered so different? Certainly because they have different methods, objects of inquiry and areas of application. These features can lead to consider the clinical and experimental inferences as different in principle. However, it is interesting to note that the kinds of inference used in both these contexts appear to be the same [Meehl (1954); Trierweiler & Stricker (1998)]. In fact, an integration between abduction and the H-D method, as the one we propose, is commonly used by the clinician as well as by the experimentalist: both individuate their objects (IO), generate different explanations about them (NHG), infer the best ones (IBE), and empirically test them (EJ). As we can see, the difference is only contingent. That is to say, it only regards the specific problems each context deals with and, consequently, the procedures of data collection and analysis, but not the way of using abduction and the H-D method.

References

Capaldi, E. J., and Proctor, R. W. (2008), 'Are theories to be evaluated in isolation or relative to alternatives? An abductive view', *American Journal of Psychology* 121 (4), pp. 617-41.

Field, A. P. (2005), *Discovering statistics using SPSS*, London, Sage Publications.

Haig, B. D. (2005), 'An abductive theory of scientific method', *Psychological Methods* 10, pp. 371–388.

— (2008a), 'Precis of "An abductive theory of scientific method"', *Journal of Clinical Psychology* 64, pp. 1019–1022.

— (2008b), 'On the permissiveness of the abductive theory of method', *Journal of Clinical Psychology* 64, pp. 1037–1045.

Johnson-Laird, P. (2006), *How we reason*, Oxford, Oxford University Press.

Josephson, J. R., and Josephson, S. G. (1994), *Abductive inference: Computation, philosophy, technology*, New York, Cambridge University Press.

[1] It is worth noting that IBE (3) is a "grey area" between CoD and CoJ. It can be considered as a form of justification, though not in the sense of comparing abstract propositions with empirical data.

Meehl, P. (1954), *Clinical versus statistical prediction. A theoretical analysis and a review of the evidence*, Geoffrey Cumberlege, Oxford University Press.
— (1957), 'When shall we use our heads instead of the formula?', *Journal of Counseling Psychology* 4, pp. 268-273.
Reber, A. S., and Reber, E. (2001), *The Penguin dictionary of psychology*, London, Penguin Books Ltd.
Romeijn, J. W. (2008), 'The all-too-flexible abductive method: ATOM's normative status', *Journal of Clinical Psychology* 64, pp. 1023–1036.
Thagard, P. (1978), 'The best explanation: Criteria for theory choice', *Journal of Philosophy* 75, pp.76–92.
— (1988), *Computational philosophy of science*, Cambridge, MIT Press.
— (1992), *Conceptual revolutions*, Princeton, Princeton University Press.
— (2000), *Coherence in thought and action*, Cambridge, MIT Press.
Trierweiler, S. J, and Stricker, G. (1998), *The scientific practice of professional psychology*, New York, Plenum Press.
Vertue, F. M. and Haig, B. D. (2008), 'An abductive perspective on clinical reasoning and case formulation', *Journal of Clinical Psychology* 64, pp. 1046-1068.

Singularismo causal

María José García Encinas
Universidad de Granada
encinas@ugr.es

El fin de esta comunicación es ofrecer una idea consistente de Causalidad Singular, según la cual la relación causal no se establece primariamente entre propiedades o entidades universales, ni sobreviene sobre, o depende en algún modo, de leyes o generalizaciones. Con este fin, analizaré diferentes definiciones y aproximaciones a la causalidad singular, en particular, la concepción de singularismo causal que emerge por contraposición directa a la teoría humeana o regularista de la causalidad, y las propuesta contemporáneas de Tooley (1984, 1987, 1999) Woodward (1990) y Ehring (2009). Argumentaré que las dos posiciones se enfrentan a importantes dificultades, y presentaré una visión intermedia. Durante la discusión emergerán y serán evaluadas diferentes ideas sobre la naturaleza de las leyes y sobre la ontología de la relación causal.

La aproximación humeana a la causalidad, i.e., la concepción de que las relaciones causales dependen del establecimiento de correspondientes universalizaciones, inspira toda una tradición generalista para el estudio y defensa de la casualidad. Según el generalismo causal, no se trata únicamente de que no podemos saber del establecimiento de hechos causales a menos que conozcamos las leyes o regularidades apropiadas, sino que se pretende un alcance metafísico fundamental: la causalidad no es primariamente una relación entre individuos, sino que la ocurrencia de hechos causales es ontológicamente dependiente de alguna relación a nivel universal. Sólo hay causalidad a nivel singular si (y porque) hay causalidad, o alguna otra relación relevante a nivel universal o general desde la que explicar, analizar o incluso reducir la causalidad.

Tal y como expone Woodward (1990: 211), según la tradición generalista los valores de verdad de los enunciados singulares causales están lógicamente determinados por los valores de verdad de los enunciados universales correspondientes —junto con los valores de verdad de los enunciados no-causales sobre los singulares implicados. (Woodward, 1990: 211). Siendo así, una primera versión, la versión más humeana y clásica, la afirmación de que la causalidad es primariamente singular consistiría esencialmente en la negación esta idea; por tanto:

> (SC1) Los valores de verdad de los enunciados causales singulares son lógicamente independientes de los valores de verdad de los enunciados universales correspondientes.

Ahora bien, el espíritu singularista causal parece estar intuitivamente bien representado por la ideas de John Duccase, el viejo filósofo singularista para la causalidad:

> "Causation is a relation which holds essentially between single, individual events though it may of course be generalized, and propositions containing kinds of events then be formulated (…) The cause of a particular event [is defined] in terms of but a single occurrence of it, and thus in no way involves the supposition that it, or any like it, ever has occurred before or ever will again. The supposition of recurrence is thus wholly irrelevant to the meaning of cause" (Duccase, 1926: 129)

En otras palabras, el singularista cree que cuando las relaciones causales se establecen a nivel singular, éstas no instancian, no sobrevienen sobre, ni dependen de ningún modo, de patrones generales o universales. La causalidad es, estrictamente, una cuestión local y particular y, por tanto, independiente de lo que pueda ser cierto antes, después, a menudo o en general, de los tipos o clases a los que los singulares del hecho causal correspondiente puedan pertenecer. Pensar que la causalidad es singular en este sentido no debería, por tanto, ser incompatible con la creencia de que (i) desde hechos causales singulares pueden hacerse correspondientes generalizaciones verdaderas, o con la creencia de que (ii) conocer las generalizaciones o las leyes apropiadas es imprescindible para comprender o conocer hechos causales singulares, o con la creencia de que (iii) para todo hecho causal existe de hecho alguna ley o generalización verdadera. Lo que el singularista causal niega es que el establecimiento de relaciones causales singulares necesite de la existencia de leyes o generalizaciones apropiadas verdaderas. Tal y como D. Ehring escribe, "causation is, thus, possible in a lawless world" (2009: 42)

Voy a argumentar que estas ideas intuitivas sobre qué es el singularismo causal no encajan con las tesis singularista humeana (SC1). En concreto, es posible proponer situaciones (e.g. Armstrong 1997: 204) que apoyan (SC1), pero que no encajan con el espíritu general del singularismo causal. Por tanto, (SC1) no representa apropiadamente la idea de que la causalidad es, en primer lugar, singular. Si estoy en lo cierto, entonces filósofos como M. Moore (2009) o Armstrong (1999) para quienes una concepción singularista es compatible con la idea de que no hay causalidad singular sin leyes siempre que el nivel singular sea respetado de algún modo en la relación —en particular, aceptando que la ontología causal es, de hecho, singular, por ejemplo, defendiendo que son sucesos concretos, o ejemplificaciones particulares de propiedades, o estados de cosas que son singulares los que conforman los hechos causales en el mundo— estarán equivocados. El problema es que estas ideas hacen del singularismo una tesis trivial; nadie, ni tan siquiera Hume, ha defendido que la causalidad ocurre exclusiva y únicamente al nivel general.

Lo que mantengo es, por tanto, que el espíritu singularista causal queda bien representado por la tesis:

> (SC) Los valores de verdad de los enunciados causales singulares son lógicamente independientes de los valores de verdad de los correspondientes enunciados nomológicos causales.

En defensa de esto, voy a considerar algunas situaciones *à la* Armstrong, como los casos de preempción y sobredeterminación de Ehring (2009), y también un argumento de Carroll (1994: 127-8) contra el universalismo causal, con el fin de defender que la apuesta por el singularismo causal debe mostrar la posibilidad de que ocurran hechos causales que son diferentes y al mismo tiempo no diferenciables por las leyes naturales que gobiernan la situación en las que estos hechos suceden. Defiendo, entonces, que el argumento de Carroll es concluyente contra el singularismo causal sólo bajo la presuposición de una teoría empirista de la leyes, pero no bajo una teoría de leyes naturales al modo de Tooley (1987), Armstrong (1983, 1997) o Drestke (1977).

Esto conlleva que el debate singularismo *versus* universalismo para la causalidad haya de dirimirse considerando tipos especiales de situaciones causales: casos de preempción causal y posibles casos de réplicas exactas, casos que los teóricos de las leyes naturales proponen precisamente en defensa de su concepción singularista. En este sentido, defiendo que los casos Armstrong y Ehring de preempción y sobredeterminación no son causales, pues no es posible que ante causas indiscernibles sólo una de ellas cause, ni que ambas lo sean del mismo efecto que habría sucedido ante la ocurrencia de tan sólo una de ellas. Aceptar lo contrario supone aceptar de nuevo, aunque inadvertidamente, una concepción humeana de la causalidad. De otro modo, lo que el análisis de estos casos muestra es que (SC2), la tesis que define el singularismo causal según Tooely y Woodward, es errónea:

> (SC2) Los valores de verdad de los enunciados causales singulares son lógicamente independientes de los valores de verdad de los correspondientes enunciados nomológicos causales *junto con los valores de verdad de los enunciados no-causales sobre los singulares implicados*.

De acuerdo con (SC2), el singularismo causal mantiene que la ocurrencia de cualquier hecho causal depende únicamente del establecimiento de una relación causal; es decir, el establecimiento de una relación causal es independiente de la ocurrencia de cualquier otra relación, hecho, propiedad u ocurrencia, incluyendo la ocurrencia de los elementos causalmente implicados. Pero las situaciones de preempción causal y sobredeterminación que podrían apoyar (SC2) son metafísicamente imposibles si se rechaza una visión humeana de la causalidad. Además, la posibilidad de réplicas exactas que el propio Tooley (1990: 1987-8) propone en apoyo de (SC2) no hace de la relación causal algo lógicamente independiente de sus propios términos. Esto significa que (SC) es la tesis que mejor representa la concepción singularista, siendo la causalidad primariamente una relación singular, ontológicamente independiente de relaciones nomológicas o generalizaciones, y que se establece entre entidades (propiedades) individuales cuya naturaleza contribuye relevantemente al hecho causal.

Referencias bibliográficas

Armstrong, D.M. (1983) *What is a Law of Nature?,* Cambridge, Cambridge University Press.
— (1997) *A World of States of Affairs,* Cambridge, Cambridge University Press.
— (1999) "The Open Door: Counterfactual versus Singularist Theories of Causation" en Howard Sankey (ed.), *Causation and Laws of Nature,* Dordrecht, Kluwer, pp. 175-185.
Carroll, J. W. (1994) *Laws of Nature,* Cambridge, Cambridge University Press.
Ducasse, J. C. (1926) "On the Nature and Observability of the Causal Relation", reimpreso en E. Sosa y M. Tooley (eds.), *Causation,* Oxford, Oxford University Press, 1993, pp. 125-36.
Ehring, D. (2009) "Abstracting Away from Preemption", *The Monist* 92(1): 41-71.
Hume, D. (1748) *Enquiry concerning Human Understanding* (diversas ediciones).
Moore, M. (2009) "The Nature of Singularist Theories of Causation", *The Monist* 92(1): 3-22.
Tooley, M. (1984) "Laws and Causal Relations", *Midwest Studies in Philosophy* 9: 93-112.
— (1987) *Causation: a Realist Approach,* Oxford, Clarendon Press.
— (1990) "Causation: Reductionism versus Realism", reimpreso en E. Sosa y M. Tooley (eds.) *Causation,* Oxford, Oxford University Press, 1993, pp. 172-92.
Woodward, J. (1990) "Supervenience and Singular Causal Statements", en D. Knowles (ed.), *Explanation and its Limits,* Cambridge, Cambridge University Press, pp. 211-46.

Cognición extendida y emulación

Ángel García Rodríguez y Francisco Calvo Garzón
Universidad de Murcia
agarcia@um.es / fjcalvo@um.es

El tema de esta comunicación es una cuestión de filosofía de la psicología y de la ciencia cognitiva: a saber, si las explicaciones de habilidades cognitivas son explicaciones que involucran procesos representacionales.

El marco que hace inteligible el planteamiento de la cuestión lo proporciona la distinción de niveles de Marr, donde hay un *explanandum* al nivel personal computacional, y dos posibles niveles explicativos subpersonales (algorítmico y de implementación). Tanto los defensores del modelo clásico [Fodor y Pylyshyn (1988)] como los del modelo conexionista [Rumelhart, McClelland *et al*. (1986)] en psicología y ciencia cognitiva defienden que la explicación de una habilidad cognitiva involucra procesos al nivel algorítmico, que representan lo que sucede en el nivel personal del *explanandum*. La explicación de una habilidad cognitiva es, pues, según estos modelos, una explicación que involucra procesos representacionales (simbólicos vs. subsimbólicos). En términos de la distinción entre vehículos y contenidos de la cognición [Hurley (1998)], lo que se está diciendo es que los vehículos explicativos de determinadas habilidades cognitivas son procesos representacionales.

Recientemente se ha propuesto un nuevo modelo para entender la cognición humana, la hipótesis de la cognición extendida, según el cual la explicación de una habilidad cognitiva involucra literalmente, no sólo sucesos neurológicos, sino también eventos corporales (no craniales) y del entorno [Clark y Chalmers (1998)]. Según un conocido ejemplo, la explicación de habilidades matemáticas cotidianas (como el cálculo de una compleja operación matemática) involucra procesos que incluyen, en parte, el uso de lápiz y papel para manipular y almacenar símbolos en el entorno, reduciendo así la complejidad del problema. Según este modelo, pues, la explicación de la cognición humana involucra procesos acoplados cerebro-cuerpo-entorno.

Los debates recientes entre defensores y detractores de la hipótesis de la cognición extendida han dado por supuesto que los procesos acoplados cerebro-cuerpo-entorno son representacionales. Como mucho, se ha discutido si las representaciones en cuestión tienen las características adecuadas para considerar a esos procesos propiamente cognitivos: es decir, si su carácter representacional es intrínseco o derivado [Adams y Aizawa (2001); Clark (2005)]. Sin embargo, la cuestión si los procesos acoplados cerebro-cuerpo-entorno involucran representaciones merece mayor atención.

La idea general de que la cognición involucra representaciones descansa sobre la intuición de que algunos procesos cognitivos son procesos hambrientos de representación [Clark y Toribio (1994)]. En otras palabras, algunas habilidades

cognitivas son habilidades *offline*, es decir, involucran objetos o estados del mundo distantes, ausentes o contrafácticos (como en la rotación de la imagen mental de un objeto ausente), de tal manera que los procesos cognitivos explicativos de esas habilidades han de ser representacionales. Es decir, dado que el objeto o estado del mundo no se halla disponible en ese momento, ha de haber algo que sustituya o esté por ese objeto en los procesos explicativos de la habilidad cognitiva en cuestión. Frente a esto, compárese lo que sucede en el caso de habilidades cognitivas *online*, en las que los objetos del mundo son manipulados directamente por el sujeto cognitivo; con lo que no hace falta apelar a nada que esté por el objeto en la explicación de la habilidad cognitiva.

Esto parecería sugerir que sólo la explicación de habilidades cognitivas *offline* requiere representaciones. Sin embargo, se ha defendido que también la explicación de habilidades cognitivas *online* requiere explicaciones en virtud de procesos representacionales. Se trata de habilidades sensorio-motoras (como restar en el tenis sin tiempo para procesar información propioceptiva sobre la posición relativa del brazo) cuya explicación involucra procesos que no forman una estructura de bucle cerrado, sino una estructura de bucle pseudo-cerrado.

La diferencia entre ambos tipos de estructura queda recogida en los siguientes diagramas (para una mayor elaboración, véase García Rodríguez y Calvo Garzón (en revisión).)

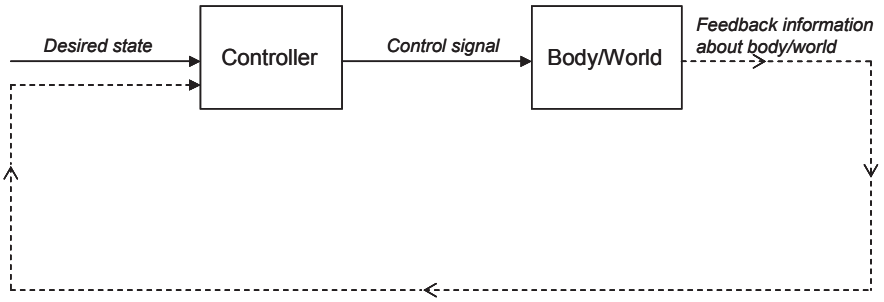


Fig. 1 Un sistema de bucle cerrado, compuesto por controlador y cuerpo/entorno.

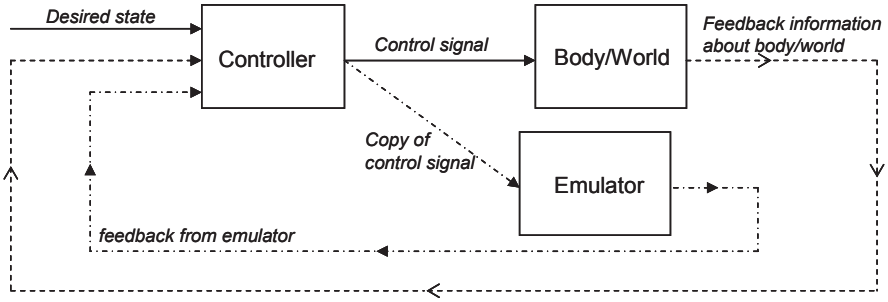


Fig. 2 Un sistema de bucle pseudo-cerrado, compuesto por controlador, cuerpo/entorno y emulador.

En ambos tipos de estructura, la información fluye de un componente, el controlador, a otro componente, el cuerpo-entorno, y vuelta, pues el controlador recibe información actualizada sobre el cuerpo-entorno, que a su vez permite el reajuste periódico de la información enviada por el controlador. (Un ejemplo clásico es un termostato, donde la función del mismo depende de la recepción de información actualizada sobre la temperatura del entorno.) La diferencia entre ambos tipos de estructura radica en que la estructura de bucle pseudo-cerrado cuenta con un componente adicional, el emulador, cuya función es permitir un desacoplamiento temporal de los procesos controlador-cuerpo-entorno. Es decir, el sistema cuenta con dos bucles: el bucle controlador-emulador y el bucle controlador-cuerpo-entorno. Ambos bucles no son totalmente independientes, pues el componente común a ambos, el controlador, envía y recibe información de ambos bucles, de tal manera que la información que fluye por un bucle afecta a la del otro. Sin embargo, los defensores del carácter especial de las estructuras de bucle pseudo-cerrado subrayan el hecho de que la información fluye por el bucle del que forma parte el emulador antes que por el bucle del que forman parte el cuerpo y el entorno, con lo que el primero está temporalmente por el segundo [Grush (2003)]. Así pues, la explicación de determinadas habilidades cognitivas (en concreto, del elemento de anticipación temporal del *explanandum*) requeriría procesos acoplados controlador-cuerpo-entorno que, al desacoplarse temporalmente, involucran representaciones [Clark y Grush (1999)].

Otra tesis mantenida por los defensores de la hipótesis de la cognición extendida, además de que las explicaciones cognitivas involucran procesos representacionales, es que los procesos cognitivos explicativos son procesos cerebrales. Así, aunque se oponen a la idea de que los procesos cognitivos no incluyen elementos no neurológicos (como objetos del entorno o implantes), al mismo tiempo afirman que el cerebro/SNC es el núcleo de la cognición [Clark (2008)]. Emerge, por tanto, una concepción de la cognición como una función de procesos centrados en el cerebro/SNC y acoplados con el cuerpo-entorno, aunque desacoplables temporalmente del cuerpo-entorno, y por tanto representacionales.

El objetivo de esta comunicación es criticar esta concepción de la cognición. En este sentido, la existencia de circuitos con emuladores en plantas es un problema. Así sucede en el caso del heliotropismo de hojas, donde las hojas de algunas plantas se reorientan durante la noche (por tanto, en ausencia del objeto) a la posición de la salida del sol, y lo siguen haciendo incluso tras tres o cuatro días sin exposición al estímulo solar. Si se asume, según la concepción de la cognición de los defensores de la hipótesis de la cognición extendida, que la cognición llega hasta donde llegan los procesos representacionales (en el sentido anterior de procesos hambrientos de representación), y que la estructura de bucle pseudo-cerrado (con emuladores) captura la arquitectura de la cognición en estos casos, entonces las plantas del ejemplo son sistemas cognitivos; con lo que no es acertado concebir la cognición como una función de procesos centrados en el cerebro/SNC.

Una posible reacción ante este ejemplo es abandonar la noción de representación. Según esto, se acepta de entrada la existencia de una similitud

arquitectónica entre la estructura de los procesos explicativos de la conducta adaptativa de las plantas del ejemplo y de determinadas habilidades sensorio-motoras humanas (el ejemplo del tenis), en cuanto que son procesos con una estructura de bucle pseudo-cerrado. Se rechaza, sin embargo, la glosa representacionalista de tales estructuras; es decir, que el bucle del que forma parte el emulador está temporalmente por el bucle del que forman parte el cuerpo y el entorno. La clave para entender las estructuras de bucle pseudo-cerrado es que el bucle del emulador es en realidad un sub-bucle del circuito en su conjunto; con lo que la idea de desacoplamiento no describe correctamente el funcionamiento del circuito. Más bien, los supuestos casos de desacoplamiento temporal son, en realidad, casos en los que los procesos explicativos tienen una complejidad causal distinta en un sistema de bucle cerrado; es decir, una complejidad causal donde unos procesos tienen lugar antes que otros. Así, el elemento de anticipación temporal del *explanandum* no se explica por un desacoplamiento temporal que invita la glosa representacionalista, sino por la existencia de distintas relaciones causales en procesos acoplados cerebro-cuerpo-entorno recogidos en la estructura de bucle cerrado. (Nótese que nada de esto implica negar la existencia de sistemas con una estructura de bucle pseudo-cerrado, o que estos puedan ser explicativamente útiles. Más bien, lo que se rechaza es que esta estructura sea sustancialmente distinta de la estructura de bucle cerrado, pues al insistir que se trata de un tipo de estructura sustancialmente distinto lo que se defiende en realidad es la idea de desacoplamiento temporal que invita la glosa representacionalista.)

Esto tiene dos consecuencias. Una, que no hay razones a partir de la arquitectura de la cognición para concluir que los procesos explicativos de determinadas habilidades humanas involucren representaciones (en el sentido de procesos hambrientos de representación). Dos, que dada la similitud arquitéctonica señalada arriba no hay razones para afirmar que los procesos cognitivos sean procesos centrados en el cerebro/SNC. Emerge, por tanto, una concepción de la cognición humana como una función de procesos acoplados cerebro-cuerpo-entorno, aunque no es el caso que la cognición *simpliciter* requiera procesos centrados en el cerebro/SNC. Dicho de otro modo, si la cognición es una función de la existencia de una determinada estructura arquitectónica en un sistema vivo, entonces puede hablarse de cognición incluso en sistemas vivos que carezcan de cerebro/SNC.

Podría replicarse que en esta concepción de la cognición se pierde lo que hace especial a la cognición humana, frente a habilidades adaptativas en general, justamente el carácter representacional de aquélla. Sin embargo, en el marco proporcionado por la distinción de niveles (personal vs. subpersonales) de Marr, esta réplica esconde una confusión entre el nivel personal de la habilidad representacional a explicar, y el nivel de los procesos explicativos de la habilidad. Negar que los procesos subpersonales estudiados por la psicología y la ciencia cognitiva involucren representaciones, en lugar de relaciones causales, no implica negar la existencia de capacidades representacionales en el caso humano. Volviendo a la distinción entre vehículos y contenidos de la cognición que se

introdujo al comienzo, negar que los vehículos de la cognición sean representacionales no implica negar las capacidades representacionales humanas.

Referencias bibliográficas

Adams, F. y Aizawa, K. (2001), 'The bounds of cognition', *Philosophical Psychology* 14, pp. 43-64.

Clark, A. (2005), 'Intrinsic content, active memory and the extended mind', *Analysis* 65, pp. 1-11.

— (2008), *Supersizing the Mind*, Oxford, OUP.

Clark, A. y Chalmers, D. (1998), 'The extended mind', *Analysis* 58, pp. 7-19.

Clark, A. y Grush, R. (1999), 'Towards a cognitive robotics', *Adaptive Behavior* 7, pp. 5-16.

Clark, A. y Toribio, J. (1994), 'Doing without representing', *Synthese* 101, pp. 401-31.

Fodor, J. y Pylyshyn, Z. (1988), 'Connectionism and cognitive architecture', *Cognition* 28, pp. 3-71.

García Rodríguez, A. y Calvo Garzón, F. (en revisión), 'Is cognition a matter of representations? Emulation, teleology, and time-keeping in biological systems'.

Grush, R. (2003), 'In defense of some 'cartesian' assumptions concerning the brain and its operation', *Biology and Philosophy* 18, pp. 53-93.

Hurley, S. (1998), 'Vehicles, contents, conceptual structure and externalism', *Analysis* 58, pp. 1-6.

Rumelhart, D. E. y Mcclelland, J.L. and the PDP Research Group (1986), *Parallel Distributed Processing - Vol. 1*, Cambridge, Mass., MIT Press.

Información a cambio de nada… ¿Es posible detectar objetos cuánticos sin mediar interacción?

Karim Gherab Martín y Carmen Sánchez Ovcharov
Harvard University / Universidad Complutense de Madrid
kgherab@fas.harvard.edu / carmen.sanchez@sek.es

Una de las líneas de demarcación entre la física clásica y la física cuántica, según la interpretación filosófica tradicional, ha consistido en afirmar que la detección de un objeto cuántico (un átomo, un electrón, un fotón, etc.) supone intercambiar al menos un cuanto de acción (constante de Planck, $\hbar$) que modifica inevitablemente el estado del objeto observado. Este hecho se conoce como "inseparabilidad objeto-dispositivo de medida" y ha sido el desencadenante de una *nueva concepción* o un *nuevo modelo de inteligibilidad* de la realidad física. No obstante, numerosos estudios (teóricos y experimentales) recientes han demostrado que estas afirmaciones filosóficas ortodoxas son incorrectas (o, al menos, inexactas) y apelan a reflexiones más sutiles. Esta comunicación ofrece algunas reflexiones filosóficas al respecto.

* * *

La interpretación filosófica ortodoxa de los procesos de interacción cuántica afirma que el intento de observación de un objeto cuántico, como un átomo, requiere el inevitable intercambio de un cuanto de acción, equivalente a un múltiplo entero de $\hbar$. La idea fundamental es que $\hbar$ simboliza la *imposibilidad* de una medida sin interacción y, con ello, destruye epistemológicamente la explicación causal determinista y anula la objetividad clásica vinculada a la noción de realidad física independiente de la observación.

Por una parte, la física clásica había interpretado los átomos (y demás objetos microscópicos) como esferas sólidas cuya posición y momento eran simultáneamente medibles con exactitud idealmente infinita. Este modelo, que interpretaba el mundo físico como un conglomerado de "bolas de billar", permitía presentar los átomos como objetos reales, con propiedades intrínsecas independientes de cualquier proceso de observación. Es decir, observar un átomo no implicaba afectar a sus propiedades cinemáticas y dinámicas. Así, la observación de un objeto microscópico resultaba ser independiente del acto de observación, que consiste en un proceso mediado por un instrumento de medida mesoscópico. Sin embargo, por otra parte, la física cuántica introdujo un elemento distorsionador de este equilibrio entre el mundo microscópico y el mundo mesoscópico, diluyendo la independencia hasta entonces concebida entre el objeto microscópico observado (por ejemplo, el átomo) y el observador (por ejemplo, un detector). El principio de indeterminación de Heisenberg impuso la imposibilidad

de medir simultáneamente la posición y momento[1] (y, análogamente, tiempo y energía) de una partícula microscópica. Y como Bohr mostró claramente en los congresos Solvay, el objeto cuántico está inexorablemente ligado al instrumento de medida: para Bohr, el objeto fenoménico es el objeto *en sí*[2] más el instrumento de medida. Desde entonces, cualquier proceso de medida se ha considerado como un proceso distorsionador del estado del objeto físico observado como consecuencia del intercambio de, al menos, un cuanto de acción. El "transporte" del cuanto de acción se ha visualizado tradicionalmente como una interacción mediada por una partícula de prueba (por ejemplo, un fotón) que es la que "extrae" la información del objeto cuántico (por ejemplo, un átomo) que queremos observar. En el caso de un fotón y un átomo, el fotón nos aportaría información acerca del estado del átomo, pero dejaría el estado de dicho átomo alterado una vez efectuada la medida.

Debido al carácter aparentemente insalvable de esta intervención, distorsionadora de las propiedades de las partículas observadas, físicos y filósofos de la física tomaron una posición epistemológica y ontológica sin precedentes, en su forma de representarse el micromundo, que se conoce como la *Interpretación Ortodoxa o de Copenhague*. Desarrollada principalemente por Bohr, de carácter instrumentalista, sostiene la idea de que la realidad es "producida" por el observador, o bien, en palabras de Heisenberg [Heisenberg (1958)] "…lo que observamos no es la naturaleza en sí, sino la naturaleza sometida a nuestro modo de interrogarla", anulando con ello la posibilidad de hablar de una realidad independiente del observador, en definitiva, independiente del proceso intervencionista de medida.

La cuestión que nos planteamos es la siguiente: ¿y si cambiara nuestra forma de "interrogar" la naturaleza; nos veríamos obligados a cambiar nuestras posiciones epistemológicas y ontológicas? Experimentos [Kwiat, Weinfurter, Herzog, Zeilinger y Kasevich (1995)] basados en un ingenioso montaje experimental propuesto en 1993 [Elitzur y Vaidman (1993)] han mostrado que es posible la detección de objetos sin mediar absorción. Este tipo de medidas son conocidas como *Interaction-Free Measurements* (IFM) y algunos también las catalogan como "contrafácticas" [Mitchison y Jozsa (2001); Penrose (1994)] porque es posible establecer una cadena causal de eventos y extraer información sobre el evento consecuente sin que se haya producido el antecedente. En general, el fenómeno que se produce es un derivado de *gedankenexperiments* más antiguos llamados "experimentos de resultado negativo" [Epstein (1995); Renninger (1953; 1960); Dicke (1981)].

Distintas versiones del montaje experimental original [Elitzur y Vaidman (1993)] que hacen uso del Efecto Zenón Cuántico [Misra y Sudarshan (1977)] han

[1] En esta presentación no entramos en el debate que ha rodeado a la propuesta hecha por Einstein, Podolski y Rosen en 1935, un *gedankenexperiment* conocido como EPR.

[2] Nos hemos permitido esta licencia de asociar a Bohr la terminología kantiana para mostrar de forma más gradual el paso de la concepción realista ingenua (modelo "bolas de billar") a la concepción instrumentalista.

conseguido mejorar la eficiencia de tal modo que los físicos experimentales [Kwiat, White, Mitchell, Nairz, Weihs, Weinfurter y Zeilinger (1999)] han sido capaces de detectar la presencia de un objeto físico sin interaccionar con él (en el sentido clásico, es decir, sin absorción ni dispersión de la partícula de prueba por parte del objeto observado), con un nivel de certeza asintóticamente cercano al 100%.

Una de las predicciones teóricas más sorprendentes que ha sido obtenida con este tipo de montajes experimentales es la posibilidad de extraer, mediante una partícula de prueba (un fotón adecuadamente polarizado) información acerca del estado superposición en el que se encuentra un átomo sin destruir o alterar el estado superpuesto del átomo [Pötting, Lee, Schmitt, Rumyantsev, Mohring y Meystre (2000); Zhou, X., Zhou, Z.W., Guo y Feldman (2001)], es decir, que es posible conocer el estado interno del átomo sin hacer colapsar la superposición a uno de sus autovectores. Es más, tampoco se vería afectado el entrelazamiento (*entanglement*) de dicho átomo con otros átomos. En resumen, se extrae información manteniendo la coherencia del átomo. Conviene describir brevemente el montaje experimental de [Pötting, Lee, Schmitt, Rumyantsev, Mohring y Meystre (2000)] con el fin de que el lector se haga una idea de en qué consiste esta aproximación "no intrusiva" a los objetos cuánticos. En aras a no complicar excesivamente la explicación, sustituiremos el rigor del tratamiento matemático de los autores por un lenguaje conceptual.

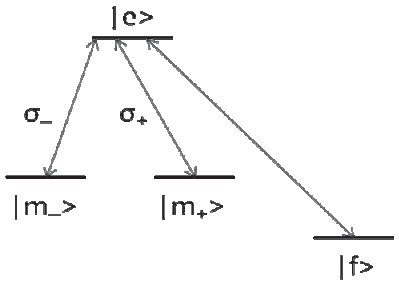


Figura 1

Tenemos un átomo que se encuentra inicialmente en una superposición cuántica de los dos estados meta-estables $|m_-\rangle$ y $|m_+\rangle$ (ver Figura 1), es decir: $|\textit{átomo}\rangle = 1/\sqrt{2}\ (|m_+\rangle + |m_-\rangle)$. En el caso de absorber un fotón, el átomo pasaría al nivel excitado $|e\rangle$, para posteriormente decaer al nivel fundamental $|g\rangle$ mediante la emisión espontánea de un nuevo fotón.

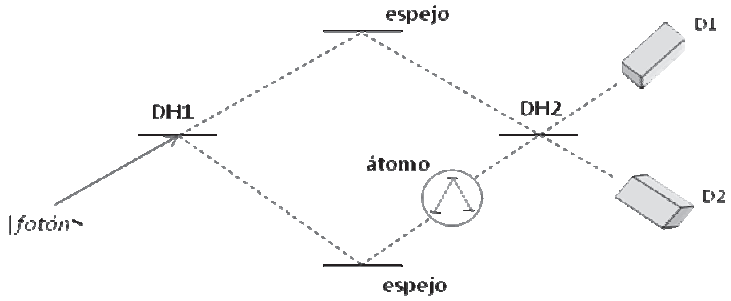


Figura 2

Colocamos el átomo dentro de un dispositivo del tipo IFM (ver Figura 2) y lanzamos hacia el dispositivo un fotón que se encuentra en una superposición de estados de polarización circular dextrógira σ_+ y polarización circular levógira σ_-, es decir[3]: $|fotón\rangle = 1/\sqrt{2}\ (|\sigma_+\rangle + |\ \sigma_-\rangle)$. Los estados de polarización σ_+ y σ_- del fotón inducen las transiciones atómicas $|m_+\rangle \rightarrow |e\rangle$ y $|m_-\rangle \rightarrow |e\rangle$, respectivamente, al interaccionar con el átomo.

En el caso de que no hubiera ningún átomo, se produciría una interferencia entre los dos caminos del dispositivo de la Figura 2, y, por consiguiente, el detector D1 indicaría la presencia del fotón con probabilidad 1. Esto implica que, si detectamos la presencia del fotón en el detector D2, entonces tenemos la certeza de que hay un átomo que obstruye el camino inferior.

Llevando a cabo una medición adecuada de la polarización lineal según el eje *x* en el detector D2, es posible demostrar que obtenemos una probabilidad 1/16 de detectar el fotón en D2 habiendo dicho fotón recorrido el camino superior. Puesto que en D2 estamos midiendo la polarización lineal del fotón según el eje *x*, esto significa que, en el caso de ser detectado, el fotón se encuentra en una superposición de los estados σ_+ y σ_-. Y puesto que el fotón ha recorrido el camino superior, el estado de superposición del átomo no ha sido afectado.

En resumen, logramos detectar y preservar la superposición del átomo porque cada componente de polarización circular (a saber, σ_+ y σ_-) del fotón lleva a cabo un IFM autónomo con el correspondiente estado meta-estable del átomo. Es decir, σ_+ detecta el estado meta-estable $|m_+\rangle$ sin que se produzca la transición $|m_+\rangle \rightarrow |e\rangle$, e, independientemente, σ_- detecta el estado meta-estable $|m_-\rangle$ sin que se produzca la transición $|m_-\rangle \rightarrow |e\rangle$. Tras llevarse a cabo ambos procesos IFM, las polarizaciones circulares se combinan para formar una polarización lineal, que es la que D2 detecta con probabilidad 1/16. Un artículo posterior [Zhou, X., Zhou, Z.W., Guo y Feldman (2001)] demostró que, haciendo uso del efecto Zenón

[3] En la práctica esto se hace lanzando un fotón de polarización lineal, digamos según el eje *x*. La polarización lineal del fotón se puede descomponer matemáticamente en término de sus polarizaciones circulares (dextrógira y levógira).

cuántico, es posible detectar y preservar la superposición del átomo con una probabilidad arbitrariamente cercana a 1.

Esta nueva posibilidad de medición no intrusiva, que ofrecen los experimentos IFM, nos obliga a revisar las afirmaciones ortodoxas que, defendidas desde hace casi un siglo y en diferentes formas epistemológicas, afirman que la constante de Planck hace imposible la objetividad vinculada a la noción de realidad física independiente de la observación. Desde el punto de vista filosófico, nuestro objetivo no es resucitar algún tipo de realismo (ingenuo, de teorías, de entidades, etc.), sino adecuar el posicionamiento epistemológico del mundo científico y filosófico a los nuevos resultados experimentales. Afirma Hacking [Hacking (1996)] que "el realismo es asunto de intervenir en el mundo, más que de representarlo en palabras y pensamiento", pues "la realidad tiene que ver más con lo que hacemos en el mundo, que con lo que pensamos acerca de él". Si por "observar" entendemos "obtener información sobre propiedades físicas", los experimentos denominados *Interaction-Free Measurements* han puesto de manifiesto que el concepto actual de "realidad" no descarta rotundamente la posibilidad de obtener información de ciertas propiedades sin alterarlas, esto es: quizás podemos *observar sin intervenir*.

Es por ello necesario, en la línea del esfuerzo de algunos físicos y filósofos de la física [Vaidman (2003; 2008); Cramer (2006); Angelo (2009)], acometer una reinterpretación de los fenómenos cuánticos desde una perspectiva filosófica renovada que tome en cuenta las sutilezas de los descubrimientos realizados en fechas recientes, trazando una nueva línea de demarcación entre el objeto microscópico observado y el observador mesoscópico que en modo alguno puede ser el tradicionalmente defendido, a saber, el intercambio de un cuanto de acción $\hbar$ mediado por una partícula de prueba.

Referencias bibliográficas

Angelo, R. M. (2009), "On the interpretative essence of the term 'interaction-free measurement': The role of entanglement", *Foundations of Physics* 39 (2), pp. 109–119.

Cramer, J.G. (2006), "A Transactional Analysis of Interaction-Free Measurements", *Foundations of Physics Letters* 19 (1), pp. 63-73.

Dicke, R. H. (1981), "Interaction-free quantum measurements –a paradox", *American Journal of Physics* 49, pp. 925-930.

Elitzur, A. y Vaidman, L. (1993), "Quantum Mechanical Interaction-Free Measurements", *Foundations of Physics* 23 (7), pp. 987-997.

Epstein, P. (1995), "The Reality Problem in Quantum Mechanics", *American Journal of Physics* 13 (3), pp. 127-136.

Hacking, I. (1996), *Representar e intervenir*, México, Paidós-UNAM, (edición original 1983).

Heisenberg, W. (1958), *Physics and Philosophy,* Nueva York, Harper.

Kwiat, P., Weinfurter, H., Herzog, T., Zeilinger, A. y Kasevich, M. A. (1995), "Interaction-Free Measurement", *Physical Review Letters* 74 (24), pp. 4763-4766.

Kwiat, P.G., White, A.G., Mitchell, J. R., Nairz, O., Weihs, G., Weinfurter, H. y Zeilinger, A. (1999), "High-Efficiency Quantum Interrogation Measurements via the Quantum Zeno Effect", *Physical Review Letters* 83 (23), 4725-4728.

Misra, B. y Sudarshan, E. C. G. (1977), "The Zeno's paradox in quantum theory", *Journal of Mathematical Physics* 18 (4), pp. 756-763.

Mitchison, G. y Jozsa, R. (2001), "Counterfactual computation", *Proceedings of The Royal Society of London A* 457, pp. 1175-1193.

Penrose, R. (1994), *Shadows of the mind*, Oxford, Oxford University Press, p. 240.

Pötting, S., Lee, E. S., Schmitt, W., Rumyantsev, I., Mohring, B. y Meystre, P. (2000), "Quantum coherence and interaction-free measurements", *Physical Review A* 62 (060101).

Renninger, M. (1953), "Zum Wellen-Korpuskel-Dualismus", *Z. Phys.* 136, p. 251.

— (1960), "Messungen ohne Stoerung des Mesobjekts", *Z. Phys.* 158, pp. 417-420.

Vaidman, L. (2003), "The Meaning of the Interaction-Free Measurements", *Foundations of Physics* 33 (3), pp. 491-510.

— (2008), "Are Interaction-free Measurements Interaction Free?", *arXiv: quant-ph/0006077v1*, <http://arxiv.org/PS_cache/quant-ph/pdf/0006/0006077v1.pdf>.

Zhou, X., Zhou, Z.W., Guo, G.C. y Feldman, M. J. (2001), "High-efficiency nondistortion quantum interrogation of atoms in quantum superpositions", *Physical Review A* 64 (020101).

Contemporary mechanistic philosophy and ecological mechanisms: the case of interspecific exploitative competition*

Rafael González del Solar
Universidad Autónoma de Barcelona
Rafael.Gonzalezd@campus.uab.es

Introduction

In the course of the last two decades there has been a noticeable increase in the philosophical literature about mechanisms. The primary goal of the so-called "new mechanistic philosophy" (Skipper y Millstein 2005) is to shed new light on some (classical) problems in the philosophy of science –such as the explanation, prediction, and unification of scientific facts– in terms of the search, discovery, modelling and testing of mechanisms. While all those studies emphasize the relevance of mechanisms to several scientific practices, as well as the interest of their philosophical examination, they are far from constituting an unified movement. In fact, despite the overlap in their epistemological accounts, contemporary mechanistic philosophers disagree on some basic issues, especially regarding the nature of mechanisms. Hence the interest of examining the adequateness of those different views on mechanisms for different scientific disciplines.

Taking lead from Skipper y Millstein's (2005) probing of the new mechanistic philosophies in the context of natural selection, I explore how the said views fare at accounting for one typical ecological mechanism, namely interspecific exploitative competition.

In order to do this, I briefly describe the most prominent features of the main new mechanistic philosophies, with a focus on ontological matters. Then, I offer a description of interspecific exploitative competition. Following the description of the chosen competitive mechanism, I assess the abovementioned philosophical views on mechanism in the context of interspecific exploitative competition. Finally, finding them wanting I turn to Mario Bunge's conception of mechanisms as *specific processes in systems* (Bunge 1964, 1997, 2004). The preliminary result of my analysis is that interspecific exploitative competition –and probably other ecological mechanisms as well– is better described as a specific processes in a system.

* The author thanks Luis Marone (ECODES, Argentina) for providing insightful comments on ecological mechanisms. The research for this presentation has been partially funded by project FI2008-01559/FISO granted to TECNOCOG research group (UAB) by the Ministerio de Ciencia e Innovación de España.

The new mechanistic philosophy

The new mechanistic philosophy comprises at least three competing views on the nature of mechanisms. These perspectives conceive of mechanisms alternatively as *complex systems*, objects or things (e.g, Glennan 2002); as *entities and activities* (Machamer, Darden and Craver. 2000; MDC); and as *unique causal chains* (Glennan 2002). The first two perspectives are much more developed than the last one, which has been offered only in schematic form. In all three perspectives, the main interest of describing a mechanism is providing an explanation for the behavior of a complex object. Besides, all three point at the importance of mechanisms in designing and interpreting experiments.

The key elements of Glennan's definition of complex-systems mechanisms are this: (a) a mechanism is for a behavior; (b) being a complex thing, composed of parts; (c) component parts interact in virtue of their organization; (d) interactions between parts reliably bring about the mechanism's behavior; and thus (e) interactions can be characterized by direct, invariant, change-relating generalizations.

According to MDC, mechanisms consist of entities that engage in activities. MDC's emphasis is in activities, which are constitutive of the transformations that produce regular changes from start to finish conditions, through a series of stages. Changes are brought about by the activities of entities in virtue of the latter's intrinsic properties, as well as of their spatiotemporal organization (location, structure, and orientation; order, rate, and duration). Regularity comes from the productive continuity between stages, which thus provides intelligibility to the connection between stages.

Glennan's mechanisms as unique causal chains are not 'robust', i.e., they do not reliably produce a behavior, so they cannot be described by means of invariant generalizations. Glennan's example of a process-mechanism is the assassination of Archduque Ferdinand triggering World War II, i.e., a unique, historically contingent fact.

Interspecific exploitative competition

The search for mechanisms is a relatively common practice in ecology, the science attempting to describe, explain and predict the abundance and distribution of organisms, as well as their interactions with the environment (including other organisms). Besides the numerous biological mechanisms relevant to ecological research, such as feeding and reproduction, there are more specific ecological mechanisms, such as competition, predation, parasite-host interactions, etc. From a philosophical point of view, ecological mechanisms and their models elicit a number of interesting questions, one of them being what ecological mechanisms are. My presentation focuses on interspecific exploitative competition, one of the ecological interactions usually invoked as a mechanism determining species abundance and diversity in ecological communities.

Individuals of different species may compete directly –by physically or chemically attacking each other (interference competition)– or indirectly –by

making some limiting resource, such as water, nutrients, shelter or light less available to each other (interspecific exploitative competition). In the pattern and process talk usual in ecology, interspecific exploitative competition is one of the processes invoked to account for certain patterns (regularities), such as a sustained decline in population numbers of organisms of species A on the one hand, and a sustained increase in organisms of the sympatric species B on the other. According to niche theory (e.g., Chase y Leibold 2003), the two basic outcomes of interspecific exploitative competition –species coexistence and extinction– are regular, and depend both on (a) the effect of changes in resource availability on the fitness of individuals of each species and (b) the per capita effect of individuals of each species on resource availability. In other words, interspecific exploitative competition is a process consisting of a sequence of stages in which individuals of species A interact with resource R changing its availability to individuals of species B, thus affecting the latter's fitness.

Mechanisms in philosophy and ecology

The previous description of interspecific exploitative competition –my choosen exemplar of an ecological mechanism– describes a process that occurs in a somewhat loose system and involves at least three heterogeneous types of interacting entities: (i) competing organisms of different species, (ii) limiting resources, and (ii) other potentially relevant items in the environment (such as climatic conditions, predators, etc.).

Interspecific exploitative competition being a process, not a thing, suffices for deeming the view of mechanisms as complex *things* not adequate for ecology. An attempt to fit interspecific exploitative competition into Glennan's component parts-interactions scheme would render a characterization of an ecological community –a system– but not that of a mechanism, and ecologists clearly distinguish between ecological communities and their mechanisms.

However, Glennan's view of mechanisms as processes does not seem to make the job either, since the author states that process-mechanisms are *unique* causal chains, a claim that contrasts with the accounts ecologists make of competitive mechanisms. These are considered types that are instantiated in a number of ways, and are modelled accordingly (see, e.g., Chase y Leibold 2003). Interspecific exploitative competition, in particular, is a mechanism type that can be realized in rather different ways depending on the particular organisms, resources and environmental conditions involved in the process. For example, while individuals of one plant species may compete for light with individuals of another shorter species by projecting their shadows on the latter, animals of species A may compete with those of species B by reducing the numbers of a shared prey to levels that are critical for the fitness of species B.

The view of *mechanisms as entities and activities* does not fit IEC either. Arguably, it is the activities (e.g., foraging) of the entities involved (organisms of at least two species, and resources) that produce changes in the system (e.g.,

population density of each species), but they do so only in virtue of their interactions. Moreover, like systems-mechanisms, mechanisms as entities and activities would imply that communities, not competition, are mechanisms.

Since the views in the so-called new mechanistic philosophy do not adequately characterize IEC, I assess one more contemporary philosophical conception that pays close attention to mechanisms, namely Bunge's systemism.

Bunge's focus is on systems, and mechanisms are just one of the aspects of systems, other being composition, structure, and environment. According to this view, a mechanism is not a complex thing, but a specific process that "makes a system tick".

In the case of interspecific exploitative competition (between two species with one resource), the individuals of competing species and the limiting resource are the component parts of the system, while its structure consists of the relations between competing species and the resource; the environment comprises all those factors (such as climatic or soil conditions) that can affect the competing species and/or the resource. Finally, the mechanism is the process by which consumption of the resource by species A changes (reduces) resource availability to species B, and thus changes (lowers) its fitness.

In sum, while the views in the new mechanistic philosophy do not fit my example of an ecological mechanism, Bunge's conception of mechanisms as processes in systems is adequate for characterizing the mechanism of interspecific exploitative competition. This suggests the interest of probing this view with other ecological mechanisms, such as predation and parasite-host interactions.

References

Bechtel, W. and R. Richardson (1993) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, Princeton University Press.

Bunge, M. (1964) "Phenomenological theories", in M. Bunge (ed.), *Critical Approaches to Science and Philosophy*. New Brunswick, NJ, Transaction, 1999, pp. 234-254.

— (1997) "Mechanism and explanation", *Philosophy of the Social Sciences* 27(4): 410-465.

— (2004) "How does it work? The search for explanatory mechanisms", *Philosophy of the Social Sciences* 34: 182-210

Chase, J. M. and M. A. Leibold (2003) *Ecological Niches. Linking Classical and Contemporary Approaches*, Chicago and London, University of Chicago Press.

Glennan, S. (2002) "Rethinking mechanistic explanation". *Philosophy of Science* 69: S342-S353.

Machamer, P., L. Darden and C. Craver (2000) "Thinking about mechanisms." *Philosophy of Science* 67: 1-25

Skipper, R. A. and R. L. Millstein (2005) "Thinking about evolutionary mechanisms: natural selection", *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Medical Sciences* 36, pp. 327-347.

La Paradoja de la Inducción de Goodman desde el Funderentismo

Ana Belén González Pérez
Universidad de Málaga
anabelengonzalezperez@yahoo.es

Introducción

¿Por qué si plantamos un naranjo, esperamos que su fruto sean naranjas? ¿Por qué no plátanos? ¿O zanahorias? Bueno, respondería cualquiera con sentido común, el fruto de los naranjos son siempre naranjas.

Asumiendo que efectivamente, todos los naranjos dan naranjas, ¿son todas las naranjas de ese color, o hay naranjas azules? Otra pregunta aparentemente absurda, respondería nuestro interlocutor exasperado. Todo el mundo sabe que las naranjas son de color naranja. A lo que responderíamos, ¿quién ha visto todas las naranjas del mundo para poder afirmar esto?

Para aclarar más la cuestión veamos un ejemplo contrario. Si preguntáramos a algunos de los presentes en esta sala a qué se dedican, lo más probable es que respondieran: a la filosofía. ¿Podría inferir entonces que todos los seres humanos son filósofos?

La estructura sintáctica es la misma, mi experiencia presente de las naranjas es que son de color naranja, así que infiero que todas las naranjas presentan ese color. Todas las personas que me rodean en este congreso son filósofos, luego todos los seres humanos son también filósofos. ¿Quién podría negar, si atendiéramos sólo a la sintaxis, que todas las naranjas son de color naranja y los seres humanos, filósofos? Aunque parece lógico, esta afirmación no resulta convincente.

En la base de estas dos preguntas está el problema de la inducción. Cómo, a partir de eventos presentes y pasados podemos inferir eventos futuros, y la segunda cuestión, cómo desde lo particular podemos inferir lo general. O, planteado de otra manera, el problema de la inducción consiste en distinguir cuándo los datos que tenemos son una muestra representativa de una característica que es generalizable a todos los eventos del mismo tipo, como el color de las naranjas, y cuándo es producto de las circunstancias, como este congreso donde, naturalmente, todos los asistentes están, de una u otra manera, relacionados con la filosofía.

De Hume a Goodman

Nelson Goodman realiza un análisis del problema de la inducción, resaltando las dificultades de fundamentación.

Inicia la cuestión exponiendo la forma en que Hume contestó a la pregunta clave: ¿Por qué elegir una predicción en lugar de otra? Esta elección entre predicciones no es una cuestión lógica, sino una cuestión práctica. Aquello que se ha producido regularmente en el pasado, lo que se ha repetido, lo que hemos experimentado, sería posible añadir, acaba por convertirse en un hábito. Así pues nuestra elección entre predicciones se basa en hábitos. (Goodman 1954, p. 64).

En el ejemplo del principio, la diferencia entre las naranjas y los filósofos. Fuera de este congreso, de los departamentos de filosofía, tenemos el hábito de no encontrar filósofos en cada ser humano, pero sí de encontrar naranjas de ese color.

Esta afirmación de Hume, aunque responde a la cuestión, no resulta del todo satisfactoria, ya que es aplicable sólo a aquellos casos que sabemos que son correctos, pero no proporciona criterios para los casos falsos. La razón es que para justificar la inducción requerimos una justificación lógica, al igual que la justificación lógica de la deducción.

¿Qué justifica la validez de una inferencia deductiva?, se pregunta Goodman, Su coherencia con las reglas de la deducción. ¿Y cómo se justifican esas reglas? Mediante su cotejo con las inferencias deductivas que se consideran válidas. Este círculo no es vicioso, sino virtuoso, afirma el autor, ya que existe un delicado reajuste entre las reglas de la deducción y la práctica de las mismas. Su mutua influencia es la que fundamenta la deducción (Goodman 1954, p. 66).

Del mismo modo el proceso de justificación de la inducción proviene de su ajuste con reglas inductivas que a su vez deberían estar de acuerdo con aquellas inferencias deductivas que consideramos válidas. De este modo Hume no parecía estar tan lejos de la respuesta como afirman sus críticos. La justificación de la inducción requiere fijarse en cómo se realizan en la práctica las inducciones. "El problema de la inducción", afirma Goodman, "no es un problema de demostración, sino el problema de definir la diferencia entre predicciones válidas e inválidas" (Goodman 1954, p. 68).

Hasta el momento los intentos de formalizar lógicamente la inducción han generado muchas dificultades y escasos éxitos. Fruto de esos intentos surge la paradoja formulada por Goodman, que ya es conocida. En resumen, viene a formular la siguiente propuesta. Supongamos que existe un predicado, *verzul*, que es aplicable a todas las cosas tal que, examinadas antes de un tiempo *t* son verdes, y las examinadas después son azules. En un tiempo *t* tendríamos la evidencia de que todas las esmeraldas son verdes, y que, al mismo tiempo, todas son *verzules*. Pero para que esto se confirmara, las esmeraldas examinadas después de *t* serían azules. (Goodman 1954, pp. 74-75).

La paradoja lógica, explica Susan Haack, consiste en que en un tiempo *t*, la relación entre "todas las esmeraldas observadas hasta este momento *t* son verdes, luego todas las esmeraldas son verdes", es confirmada del mismo modo que "todas las esmeraldas observadas hasta este momento *t* son *verzules*, luego todas las esmeraldas son *verzules*". No existe diferencia lógica alguna entre 'verde' y 'verzul' (Haack 2003, p. 84).

Esta paradoja trata pues con los dos aspectos de la inducción. Uno de ellos es la proyección de eventos pasados a eventos presentes y futuros, y el otro aspecto es la generalización de eventos presentes y pasados a eventos pasados, presentes y futuros.

Malas interpretaciones

Rami Israel señala dos posibles interpretaciones de la paradoja de Goodman. Una de ellas errónea y además, afirma el autor, ampliamente extendida. Esta mala interpretación caracteriza la paradoja de la siguiente forma.

En un tiempo *t* las cosas son verdes, la hierba, las esmeraldas, las hojas de los árboles. El predicado verzul implica que a partir de un tiempo *t,* las cosas han cambiado de color, y la hierba, las esmeraldas, se han transformado en cosas azules. En esta versión se elimina el elemento 'ser observadas'. Independientemente de la observación, las cosas son verdes antes de *t* y azules después de *t.* Esta interpretación priva a la paradoja de su fuerza, ya que habría que explicar por qué las esmeraldas se transforman. Si no hay ninguna causa de transformación, la paradoja no tiene sentido (Israel 2004, p. 336).

La otra interpretación señala con gran potencia el problema de la inducción. Si tenemos un tarro de caramelos y hasta un tiempo *t* todos los caramelos que hemos extraído del tarro eran verdes, ¿cómo sabemos que los caramelos que quedan son también verdes? ¿Cómo podemos distinguir que no ha sido el azar el que ha causado que hasta el momento hayamos extraído los caramelos verdes, pero que los que quedan en el tarro son todos azules? En este caso la paradoja prevalece en toda su fuerza. La pregunta que se hace es cómo distinguimos en una inferencia deductiva las inferencias casuales "todas las esmeraldas son *verzules*" de las inferencias con fuerza de ley "todas las esmeraldas son verdes" (Israel 2004, p. 338).

La propuesta funderentista

El funderentismo es una propuesta epistemológica formulada por Susan Haack. Fundamenta la posibilidad del conocimiento en un sistema que sintetiza el coherentismo y el fundamentalismo. Esto implica que el conocimiento es posible gracias a la colaboración entre las reglas lógicas del discurso y los aportes de la experiencia (Haack 1993).

Desde el funderentismo, Haack analiza la paradoja y aporta una solución que toma tanto de las propuestas de Goodman como de Quine. Goodman, fiel a la propuesta de Hume acerca de los hábitos, la evoluciona, y responde a su propia paradoja aludiendo a razones socio-históricas, afirma Haack. Es el uso científico el que legitima las inferencias que pueden ser proyectadas. También Quine propone una resolución a la paradoja, cuando afirma que sólo los predicados de categorías naturales son proyectables. Ninguna de las dos respuestas satisface del todo a la autora. La primera porque denota un conservadurismo científico. ¿Acaso sólo un científico puede realizar inferencias inductivas válidas? En el caso de Quine, la pregunta, irónica, sería ¿constituyen las cosas verdes una categoría natural? (Haack 2003, p. 86).

Sin embargo, añade Haack, ambos tienen razón en una cosa. Tanto las razones socio-históricas como las categorías son parte de la trama del conocimiento que se ha ido tejiendo a lo largo de los años. Esta trama no sólo incluye, en el caso de las esmeraldas, la experiencia de que todas las esmeraldas vistas hasta ahora son verdes, sino una completa trama de indicios acerca de la percepción, la óptica, la composición de las gemas, la clasificación de las esmeraldas, etc., que constituyen bases sobre las que proyectar la afirmación de que todas las esmeraldas son verdes y no *verzules*.

Haack utiliza la metáfora del crucigrama para fundamenta la labor del conocimiento, no sólo científico sino cualquier tarea epistemológica humana. Una prueba consolidada se entrelaza con una pista apenas segura, y cada nueva afirmación, cada nueva inferencia inductiva es confirmada por las demás palabras, las demás entradas del crucigrama, que son a su vez confirmadas por la nueva inferencia. En este caso el color verde de las esmeraldas se entrelaza con esas características ópticas, químicas, clasificaciones, etc., dando una validez mayor tanto a la proyección de esa cualidad desde la experiencia pasada a la futura como a la generalización.

Además, el vocabulario que plantea Goodman, *verzul*, *azurde*, aunque desde el punto de vista lógico no sea inapropiado, sería difícil de aprender en la práctica, y sobre todo de utilizar. Después de un tiempo *t* los científicos descubrirían que las esmeraldas no son *verzules* sino *azurdes*, y que algo en su percepción del *vercolor* estaba rematadamente mal, así que tarde o temprano sería reemplazado por el actual. (Haack 2003, p. 86).

En una lectura apresurada, este último párrafo podría interpretarse como una visión errónea de la paradoja por parte de Haack, pero la autora no alude a que los científicos esperen que las esmeraldas cambien de color al aceptar el vocabulario del *vercolor*, sino que existan, en un momento determinado, singularidades, esmeraldas que aparezcan con un color distinto. Sin embargo, resultaría mucho más práctico decir que hay esmeraldas verdes y algunas azules, que decir que las esmeraldas son *verzules*.

Para Haack fundamentar la inducción en esta trama del conocimiento no implica saber siempre distinguir entre una inferencia inductiva válida y una casual. Hay errores, porque nuestro conocimiento es falible. Pero como ella misma afirma, ningún epistemólogo escéptico ha utilizado jamás la paradoja de Goodman para negar la posibilidad del conocimiento. (Haack 2003, p. 86).

Conclusión

La paradoja de Goodman pone de relieve la dificultad que existe para justificar la inducción, y lleva a la cuestión fundamental de qué se entiende por confirmación. Hasta el momento, no ha podido ser resuelta sólo con la ayuda de herramientas lógicas, y desde el punto de vista del funderentismo, no puede serlo. La posibilidad del conocimiento se da gracias a la colaboración entre las leyes lógicas y el amplio corpus de conocimiento que se ha ido entrelazando a medida que la investigación humana avanza. Aunque no existan normas absolutamente

irrefutables que nos permitan fundamentar las inferencias inductivas, existe una trama que permite sostenerlas como inferencias básicas que posibilitan la adquisición de conocimiento. La trama formada por la experiencia humana.

Referencias bibliográficas

Goodman, N. (1954), *Fact, Fiction and Forecast,* London, University of London, Athlone Press.

Haack, S. (2003), *Defending Science within reason*, Amherst, New York, Prometheus Books.

— (1993), *Evidence and Inquiry*, Oxford, Blackwell.

— (1982), *Filosofía de las lógicas*, Madrid, Ediciones Cátedra S.A.

Israel, R. (2004), 'Two Interpretetions of 'grue' - or how to misunderstand the new riddle of induction', *Analysis* 64 (4), pp. 335-39.

Rescher, N. (1980), *Induction,* Oxford, Basil Blackwell Publisher.

Swinburne, R., ed., (1976), *La justificación del razonamiento inductivo,* Madrid, Alianza Universidad.

El panrelacionismo rortiano en la intepretación de los objetos científicos: una perspectiva para la dualidad onda-partícula

Nalliely Hernández Cornejo
Universidad Complutense de Madrid
nallie3112@hotmail.com

Richard Rorty es un filósofo en el que convergen la tradición analítica, la tradición continental y el pragmatismo americano. Durante la segunda mitad del siglo XX, el filósofo norteamericano desarrolla una propuesta donde la concepción representacionista del conocimiento y sus premisas son el centro de su crítica, así como el uso y reinterpretación de algunas tesis de estas líneas de pensamiento son su instrumento principal. A través de esta fusión, Rorty conforma una propuesta neo-pragmatista, ecléctica y controvertida, que pretende sobre todo superar los problemas clásicos de la epistemología.

De acuerdo con Rorty, desde el siglo XVII, se ha consolidado la concepción de que el conocimiento es el proceso en el que intentamos representar la realidad, *tal como es*. Sin embargo, dicha representación se ve constantemente impedida porque existe una barrera entre nosotros y ella al haber: "… un velo de apariencias, producido por la interacción entre sujeto y objeto, entre la constitución de nuestros órganos sensoriales o nuestras mentes y la manera en que las cosas son en sí mismas." [Rorty (1997 [1994]), p. 47].

En contraposición, Rorty comparte con Peirce y Dewey la idea de que toda indagación parte de la necesidad de satisfacer las distintas necesidades del individuo [Rorty (1996 [1991]), p. 132]. Por tanto, todo objeto de conocimiento es producto de la investigación realizada con un propósito dado. Un objeto posee un conjunto de propiedades o rasgos que son justamente el resultado final de la investigación, es decir, dicho conjunto implica la posibilidad de hacer algo en relación al problema que suscitó la indagación. De esta forma, el pragmatismo pretende sustituir la noción de conocimiento como representación, por la de un instrumento para manipular los objetos [Rorty (1997 [1994]), p. 47].

Si además retomamos la tesis que Rorty hereda de Quine sobre el holismo epistemológico; la doctrina que propugna la concepción de que nuestro conocimiento es un todo distinto de la suma de las partes que lo componen entonces, no puede haber una parte de él al margen del resto. Todos los objetos resultantes de la investigación son parte de un sistema conceptual cuyos elementos están interrelacionados unos con otros.

De estas concepciones, se deduce que el pragmatismo rortiano requiere dos premisas:

1. Que el conocimiento sobre x, es equivalente a afirmar que es posible hacer algo con ese x, es decir, que es posible poner a x en relación con algo más.
2. Que dicha relación es extrínseca a x.

Estas dos premisas evitan, por un lado, la idea de que el conocimiento pueda ser contemplación o reproducción de una realidad ajena a la interacción. Por otro, afirman que no puede haber un rasgo de un objeto que no sea relacional, así como no hay una cosa que sea la naturaleza intrínseca, la esencia del objeto. La distinción entre intrínseco y extrínseco se disuelve y no hay descripción de un objeto que vaya más allá de su relación con las necesidades humanas. Así, todo lo que conocemos de un objeto es una trama de relaciones con otros objetos, sin un núcleo que no sea relacional [Rorty (1997 [1994]), p. 52]. Adicionalmente, toda relación es referida, implícita o explícitamente, en las oraciones que lo describen.

El discurso rortiano pretende evitar lo no relacional, el conocimiento directo: "…una descripción del objeto consigo mismo, de la identidad con su propia esencia" [Rorty (1997 [1994]), p. 53]. Para él, todas estas ideas recrean inevitablemente la cosa en sí kantiana, que no sólo es metafísica en el sentido más llano de la palabra, sino oscura e inaccesible. Sin embargo, no es que cualquier descripción valga igual, por el hecho de no haber propiedades inherentes a los objetos. Algunas descripciones serán mejores que otras, pero esto es debido a que resultan ser mejores instrumentos para un propósito dado. Al mismo tiempo, los seres humanos pueden tener diversidad de necesidades y propósitos, ninguno de los cuales resulta ser "el propósito primordial". Para el pragmatismo, el conocimiento como resultado de la investigación tiene como objetivo constituir creencias útiles [Rorty (1997 [1994]), p. 53].

Así, el argumento antiesencialista radica en decir que todo lo que sabemos de un objeto son las oraciones acerca de él que se consideran verdaderas, pero una oración no es más que establecer una relación del objeto con otros objetos. Por el contrario, el esencialismo suele defender que existe una parte del conocimiento que no puede ser captada por el lenguaje y radica en el contacto directo con el objeto, en sus poderes causales "intrínsecos" (como tocar el objeto). Pero, de acuerdo con Rorty: "…no intimamos más con la mesa, no nos arrimamos a su naturaleza intrínseca cuando la golpeamos, la miramos o hablamos de ella. Todo lo que hace el golpearla o descomponerla en átomos es permitirnos relacionarla con unas cuantas cosas más. Ello no nos conduce del lenguaje al hecho o de la apariencia a la realidad o de una relación remota y no interesada a una relación más íntima e intensa" [Rorty (1997 [1994]), p. 57].

Sin embargo, no es que el pragmatista ponga en duda la existencia del mundo independiente de él. Lo que no cree es que la pregunta acerca de qué pueden ser estos objetos, además de sus relaciones con otros objetos, tenga algún sentido o utilidad. Lo que un objeto pueda ser más allá de nuestras descripciones es la cosa incognoscible que resulta inexpresable y vacía. Una afirmación que resulta tautológica, pues definimos algo incognoscible porque, claro, está más allá de nuestro conocimiento [Rorty (1997 [1994]), p. 60].

En resumen:

> "Los antiesencialistas proponen rechazar todas las cuestiones acerca de dónde termina una cosa y dónde comienzan sus relaciones, todas las cuestiones relativas a dónde comienza su naturaleza intrínseca y dónde empiezan sus relaciones, todas las cuestiones acerca de dónde concluye el núcleo esencial y comienza la periferia accidental" [Rorty (1997 [1994]), p. 59].

La insistencia en tales distinciones esencialistas es constantemente preservada en el discurso del conocimiento, particularmente en el de la ciencia. A menudo se defiende la creencia de que la ciencia, y en particular, la física nos permite acceder al mundo esencial, intrínseco, no relacional, al margen de necesidades y propósitos. Por el contrario, aquí sostengo que algunos ejemplos de la ciencia contemporánea permiten una lectura plausible y consistente de estas tesis rortianas.

El principio de complementariedad

Este principio surge como base interpretativa del formalismo cuántico en el año 1927, siendo resultado de las reflexiones en torno a cómo traducir las estructuras matemáticas, que daban cuenta de los fenómenos atómicos, en una explicación física consistente. Bohr y Heisenberg debatieron el asunto con particular vehemencia a finales de 1926. Después de un tiempo y de forma separada, elaboraron el principio de complementariedad y de incertidumbre, respectivamente, los cuales conformaron la base de lo que se denominó la interpretación de Copenhague.

De acuerdo con Bohr, clásicamente, la interacción entre un fenómeno o un objeto estudiado y el instrumento de medición resulta despreciable. Sin embargo, en mecánica cuántica, debido a la indivisibilidad del cuanto de acción, dicha interacción es parte inseparable del fenómeno. Por otro lado, los instrumentos de medida están definidos en términos clásicos. Esto en la perspectiva de Bohr implica que requerimos de usar las concepciones clásicas pero elaborando un nuevo marco lógico para su uso.

Así, tomando en cuenta la inevitable interacción y el rescate de las categorías clásicas de onda y partícula, Bohr elabora el principio de complementariedad. En él se establece la dualidad onda-partícula para interpretar los fenómenos en la teoría cuántica, de tal manera, que se eviten los inconvenientes de las nociones clásicas. Todos los hechos sobre la luz y la materia pueden ser explicados en términos de uno de estos dos conceptos, pero no de los dos simultáneamente, dado que tienen propiedades excluyentes. Así, algunos sucesos se explican haciendo uso de la noción corpuscular y otros de la ondulatoria, dependiendo del contexto experimental.

De acuerdo con el principio, no existe un sistema único de descripción que sea compatible con todos los hechos, es decir, la utilización de un conjunto de conceptos clásicos (onda o partícula) en la descripción de un sistema cuántico excluye la utilización de otro conjunto que es "complementario". Sin embargo,

ambos sirven para resumir, sintetizar y unificar resultados experimentales de una forma económica.

La idea de la complementariedad pretende resolver la aparente incompatibilidad entre los aspectos fenoménicos de la onda y el corpúsculo en los procesos atómicos. La exclusión mutua y la interacción como parte del fenómeno son centrales en su interpretación. Con ella, Bohr pretende trasladar la consistencia del formalismo matemático cuántico, carente de contradicciones internas, al plano del lenguaje ordinario modificando convenientemente su alcance. Lo central es que el dispositivo experimental completo sirve para definir en términos clásicos las condiciones bajo las cuales aparece el fenómeno.

Particularmente, como el postulado cuántico impone una limitación en determinar la posición en relación al momento. El electrón solo muestra su carácter corpuscular al ser observada su posición, pero se interfiere su propiedad dinámica del momento destruyendo el fenómeno de interferencias que causa la onda. Al observar su impulso, su localización es indeterminada, sólo se presenta la propiedad ondulatoria. Esto sintetiza su comportamiento dual. Toda propiedad emerge de la relación con el instrumento de medición.

Visiones rortianas y complementariedad

En primer lugar, las propiedades de los objetos cuánticos aparecen como resultado de una inevitable interacción, en congruencia con una visión pragmatista del conocimiento y evitando la metáfora visual y pasiva del conocimiento. La descripción del sistema cuántico sólo surge como resultado directo de la posibilidad de hacer algo con el sistema, del dispositivo experimental que usamos. Es decir, en función de este marco de descripción, los estados cuánticos se definen de esta forma como una relación entre aparato-sistema que, dependiendo de su configuración, determinará el comportamiento a observar, ondulatorio o corpuscular. Se trata, en opinión de Bohr, de una hipótesis objetiva requerida por el formalismo donde el estado físico está definido por una relación más que por una propiedad. Un sistema no *posee* ninguna propiedad más allá de las que derivan de su estado de descripción [Bohr(1964), p. 83]. Por tanto, toda propiedad es relacional y vemos satisfecha la primera premisa rortiana.

En segundo lugar, dicho carácter relacional elimina el carácter intrínseco de cualquier propiedad atómica. Las características corpusculares u ondulatorias, al depender del contexto experimental, no pueden ser tomadas como esencias del objeto, no hay forma de disociar la relación de aquello que es relacionado, pues la identidad del objeto está determinada por dicha relación y se transforma junto con ella. Ni la onda ni la partícula constituyen la "naturaleza intrínseca" de ningún objeto cuántico, la distinción entre extrínseco e intrínseco desaparece, tal y como lo establece la segunda premisa rortiana antes descrita. El propio Bohr afirmaba que la teoría cuántica nos colocaba en un nuevo contexto donde, al describir los fenómenos, nuestro propósito no es revelar su esencia misma sino establecer sólo, y en la medida de lo posible, relaciones entre los múltiples aspectos de nuestra experiencia.

Asimismo, la descripción en este marco resulta mejor porque proporciona consistencia a la explicación física de los fenómenos y rescata el uso de las categorías clásicas, dos objetivos centrales para Bohr. De la misma forma, la interpretación, a pesar, de su controversia, resultó práctica y eficiente. Por lo tanto, podemos decir que la interpretación es mejor de acuerdo con determinados objetivos, así como, útil.

Por otro lado, en esta interpretación, aquello que no sea observable y por tanto, lo que no está en el lenguaje cuántico no puede ser formulado en el mundo microfísico, al igual que en el pensamiento rortiano, por estar más allá de las posibilidades de descripción, esto es, de nuestro conocimiento. Sin embargo, no es que la interpretación de Copenhague niegue la existencia del mundo atómico independiente, pero la pregunta del comportamiento atómico al margen del sistema de descripción cuántico, al margen de las relaciones establecidas en él, carece de sentido. Nuestro conocimiento de un objeto atómico está constituido por el lenguaje cuántico, por las oraciones consideradas verdaderas acerca de ondas y partículas. De tal forma, que preguntar por algo más allá de este lenguaje, es de nuevo preguntar por lo que resulta metafísico, en consecuencia, inaccesible y fútil.

Sin embargo, no se afirma con ello, que el formalismo cuántico o alguna de sus interpretaciones sean definitivos, o que nunca se podrán alcanzar otras descripciones que incluyan aspectos de los fenómenos hasta ahora desconocidos. Las teorías y descripciones pueden variar a cualquier nivel y de formas impredecibles, el pragmatismo rortiano no impone condiciones de posibilidad o limitaciones preconcebidas para el conocimiento. Sólo quiere decir que en el marco de la complementariedad, en los criterios de un discurso actual, no tiene sentido formular propiedades que no son formulables en dicho lenguaje, que no representan una diferencia que se manifieste en la práctica.

Podemos concluir, que en el dualismo onda-corpúsculo se evita la cuestión de distinguir entre el núcleo esencial y la periferia accidental, mostrando que la física no nos da acceso a un mundo intrínseco o no relacional, resultando ser dicho esencialismo sólo un resquicio metafísico de viejas concepciones. De tal forma que la complementariedad permite una lectura rortiana de un panrelacionismo y antiesencialismo, plausible y consistente, de los objetos cuánticos. Por tanto, es estimable para este caso afirmar que, tal y como dice Rorty, la cosa en sí kantiana no sólo es metafísica, sino oscura e inaccesible.

Referencias bibliográficas

Bohr, N. (1964), *Física atómica y conocimiento humano,* Madrid, Aguilar.

— (1970), *Nuevos ensayos sobre física atómica y conocimiento humano: 1958-1962,* Madrid, Aguilar.

— (1988), *La teoría atómica y la descripción de la Naturaleza: cuatro ensayos precedidos de una introducción*, Madrid, Alianza.

Rorty, R. (1997 [1994]), *¿Esperanza o conocimiento?: Una introducción al pragmatismo*, México, Fondo de Cultura Económica, 1997.

— (1996 [1991]), *Objetividad, relativismo y verdad*, Barcelona, Paidós.

¿De dónde vienen las poblaciones? Modelos, representación y política en Ecología de Poblaciones

Andoni Ibarra y Jon Larrañaga
Universidad del País Vasco / Euskal Herria Unibertsitatea
andoni.ibarra@ehu.es

Durante mucho tiempo la representación mediante modelos ha sido vista como un proceso de construcción de una imagen más o menos adecuada de un sistema u objeto ya dado. Esta concepción de representación es cuestionada por nuevos enfoques en los que lo representado no se presenta como algo ya dado sino como algo constituido en el proceso mismo de representación. La representación se concibe así como un proceso en el que se describe el objeto de estudio y simultáneamente se interviene sobre él [Ibarra y Mormann (2006)]. En esta comunicación pretendemos avanzar algo en la clarificación de esta cuestión: ¿cómo es posible describir algo que, simultáneamente, se contribuye a modificar? En particular, queremos dar cuenta de cómo diferentes actores con capacidad de agencia diversa constituyen una parte de la "naturaleza" al tiempo que pretenden analizarla de manera objetiva y supuestamente externa a ella. Para ello consideraremos algunas prácticas de modelización en la Ecología de Poblaciones.

En esta disciplina científica es habitual el uso de modelos estocásticos para evaluar las probabilidades de supervivencia de poblaciones en riesgo de extinción. Los modelos poblacionales estocásticos, a diferencia de los modelos deterministas, incluyen la influencia de fenómenos aleatorios (tormentas, sequías, incendios…). En consecuencia, al contrario de lo que ocurre con los modelos deterministas, en los que a menudo representan las poblaciones como entes cuyas dinámicas tienden a situaciones estacionarias o de equilibrio y en los que siempre, si se parte de una misma situación inicial, se llega a un mismo estado final [Jackson *et al.* (2009)], en los modelos estocásticos cada vez que se ejecuta el modelo se producen diferentes resultados, aun partiendo de los mismos valores iniciales [Hilborn y Mangel (1997), p. 32] y, lo que es más importante, estos modelos suelen predecir altibajos considerables y, frecuentemente, la disminución de la población hasta su extinción en pocas generaciones [Lande *et al.* (2003), pp. 25-52]. Es decir, los modelos estocásticos muestran, a diferencia de los modelos deterministas, que para cualquier población la extinción es un hecho inevitable, más probable cuanto menor sea la población, por lo que dan lugar a una concepción de las poblaciones en la que las ideas de equilibrio y de permanencia son sustituidas por una clara conciencia sobre la fragilidad de las poblaciones y la inevitabilidad de las extinciones.

Los modelos estocásticos han introducido conceptos como el de *tiempo hasta la extinción*, que es el tiempo medio que se espera durará una población hasta su

desaparición, o el de *población mínima viable*, que es una estimación de tamaño mínimo que debería tener una población para que su probabilidad de persistir durante un periodo de tiempo determinado (p. ej. 100 años) supere cierto valor (p.ej. 90%) [Ricklefs y Miller (1999), p. 372]. Estos conceptos se han convertido en la base de los *análisis de viabilidad poblacional*, que son procedimientos para estimar el riesgo de extinción de las poblaciones y sirven, entre otras cosas, para evaluar los planes de conservación de poblaciones en riesgo con el fin de calcular superficies mínimas de hábitat necesarias para mantener poblaciones viables [Lande *et al.* (2003), pp. 103-104].

El aspecto que queremos enfatizar aquí es que los modelos estocásticos no tienen efectos por sí mismos en relación a una pretendida descripción objetiva acerca del estado de una población, sino que han de entenderse más bien como instrumentos que producen una perspectiva común que permite situar en un único escenario la acción –con frecuencia contrapuesta- de diversos actores en una situación de incertidumbre. Habitualmente tiende a pensarse que las poblaciones tienen sus propias dinámicas y que, por ejemplo, la política de gestión ambiental aplica los modelos teóricos que garantizan de la manera más adecuada posible la conservación de la población. En esta perspectiva, por tanto, los modelos serían guías para la toma de decisiones en materia de política pública. Un análisis más detenido de las prácticas de modelización en la Ecología arroja, sin embargo, una imagen distinta sobre ellas: las acciones de gestión ambiental –de muy diverso tipo- son constitutivas de la naturaleza misma de la población; sin ellas no es posible producir los cambios que garanticen a las poblaciones una –limitada- perdurabilidad. Los actores involucrados en esas acciones participan en la constitución de las poblaciones y en su dinámica. ¿Cómo es posible que los actores describan mediante modelos estocásticos la dinámica de una población determinada y, simultáneamente, actúen sobre ella transformándola, constituyéndola?

Para introducir el alcance y las razones de la constitución de las poblaciones mediante modelos de representación en la Ecología de Poblaciones, consideraremos brevemente dos aspectos relevantes de las representaciones científicas: (i) la relación entre la descripción y la acción que procuran, y (ii) la naturaleza de la articulación necesaria en la constitución de lo representado. Para ello presentaremos la polémica habida en torno a las estrategias para la conservación de la subespecie norteña del búho moteado (*Strix occidentalis caurina*) en la costa oeste de los Estados Unidos.

Este búho, del que se estima que sobreviven de 2500 a 3000 parejas, anida únicamente en bosques de coníferas maduros (de más de 80 años) situados a altitudes medias; un tipo de bosque muy rentable para la industria maderera [Doak (1989)]. Sin embargo, la constatación de que sus poblaciones se encuentran en declive llevó a implementar un plan de conservación que prohíbe la tala de grandes superficies de bosque *a priori* adecuadas para este búho con el fin de impedir su desaparición. Ello ha dado lugar a una larga polémica en la que se han visto implicados, entre otros, la industria maderera, diversos organismos gubernamentales, organizaciones conservacionistas e investigadores de diversos

ámbitos académicos que han tratado de evaluar las medidas adoptadas, tanto en relación a su influencia en las poblaciones de búho moteado, como en la de las consecuencias en la economía de las regiones afectadas.

En relación a i). Los modelos teóricos utilizados en la polémica producen la aserción de que la población de búho moteado se extingue porque se dan ciertas condiciones específicas en la costa oeste. Es importante remarcar que esta aserción no es una mera descripción complementaria a otras relativas a la naturaleza de la población. Y a la inversa: la población a la que se refiere la aserción carece de significación real sin, justamente, esa aserción que promueve la realización de determinadas acciones. Dicho de otro modo: el peligro de extinción del búho moteado no puede concebirse como hecho científico a menos que se encaje ese hecho en un escenario de representación que incorpore ya la descripción de las *situaciones reales* que pueden causar la extinción de la población. Causas –y las series de líneas de acción subsiguientes- que pueden haberse producido o no. En esta condición de la incorporación en los modelos estocásticos de situaciones y acciones por producirse, esto es, en la condición de una descripción de una población no concebida en ellos como preexistente, radica el carácter constitutivo de los modelos. Este es el primer aspecto de los modelos estocásticos de la Ecología de Poblaciones que queríamos apuntar: la interacción entre la descripción de una población que los modelos ofrecen y la acción que ellos procuran.

En relación a ii). El segundo aspecto a resaltar sobre la naturaleza constitutiva de los modelos está estrechamente relacionado con el aspecto anterior. Hace referencia a la compleja y heterogénea articulación de elementos intervinientes en la constitución de una población. En la polémica que nos ocupa, por ejemplo, diversas organizaciones conservacionistas y biólogos han cuestionado la eficacia de las medidas adoptadas, exigiendo la protección de mayores extensiones de bosque. Se ha argumentado, por ejemplo, que el plan de conservación propuesto por el gobierno ocasionaría un descenso dramático de la población de búhos pues no facilitaría la colonización de nuevos fragmentos de hábitat [Lambert *et al.* (1994)] o se ha llegado a afirmar que, debido a la dificultad de estimar qué superficie de bosques maduros garantizaría la pervivencia del búho moteado, debería suspenderse totalmente la explotación de este tipo de bosques [Doak (1989)]. Sin embargo, otros actores han argumentado que la prohibición de talar cientos de hectáreas de bosque ha sido y será la causa de una notable disminución en la cantidad de puestos de trabajo relacionados con la explotación maderera y, en consecuencia, tendrá un severo impacto social y económico en las zonas rurales afectadas [Waters *et al.* (1994)]. A ello se ha objetado con razones diferentes: que la pérdida de puestos de trabajo obedece a otros motivos, que a largo plazo la preservación de los bosques producirá mayor número de puestos de trabajo [Freudenburg et al. 1998] o que los beneficios económicos de la preservación de la especie serán superiores a los costos que conlleva [Loomis y White (1996), p. 198].

Estos breves trazos de los argumentos enfrentados en la polémica son indicativos de la diversidad de situaciones que pueden producirse a partir de las diferentes articulaciones de series de acciones de los actores que hacen uso de los

modelos en la descripción de la población de búhos. Esta descripción sigue una cadena –no lineal- de *actualizaciones* que configura finalmente a la población como una realidad objetiva. En primer lugar se la representa como una población con un determinado tiempo hasta la extinción; se calculará luego para ella la población mínima viable y el dato se integrará de nuevo en el tiempo de extinción; una población para la que construirá un lenguaje específico para describirla, en el que, finalmente, la noción de viabilidad poblacional, una simple fórmula, resulta calculable. La población del búho moteado de la costa oeste ha quedado de este modo objetivamente constituida, en el sentido de que esa fórmula permite describir el peligro real de su extinción.

En esta comunicación hemos pretendido apuntar dos aspectos de las prácticas representacionales modelísticas en la Ecología de Poblaciones: la necesaria interacción en los modelos utilizados entre la descripción y la acción, y el complejo mecanismo de ajuste –no directo- entre el modelo y lo representado, mecanismo en el que intervienen actores diversos (científicos, instituciones políticas, organismos conservacionistas, empresas) que utilizan los modelos en configuraciones de acciones muy diversas que permiten, finalmente, *actualizar* la población, dotarla de realidad objetiva. Ambos aspectos son relevantes para una consideración del carácter constitutivo de las representaciones científicas.

Referencias bibliográficas

Doak, D. (1989), 'Spotted owls and old growth jogging in the pacific northwest', *Conservation Biology* 3(4), pp. 389-396.

Freudenburg, W. R., Wilson, L. J. y O'Leary, D. (1998), 'Forty years of spotted owls? A longitudinal analysis of jogging industry job losses', *Sociological Perspectives* 41(1), pp. 1-26.

Hilborn, R. y Mangel, M. (1997), *The ecological detective: confronting models with data*, Princeton, Princeton University Press.

Ibarra, A. y Mormann, Th. (2006), 'Scientific theories as intervening representations', *Theoria* 21(1), nº 55, pp. 21-38.

Jackson, L. J., Trebitz A. S. y Cottingham, K.L. (2000), 'An introduction to the practice of ecological modeling', *BioScience* 50(8), pp. 694-706.

Lamberson, R., Noon, B. R., Voss, C. y McKelvey, K. S. (1994), 'Reserve design for territorial species: the effects of patch size and spacing on the viability of the Northern spotted owl', *Conservation Biology* 8(1), pp. 185-195.

Lande, R., Engen, s. y Saether, B. (2003), *Stochastic population dynamics in ecology and conservation*, Oxford, Oxford University Press.

Loomis, J. B., y White, D. S. (1996), 'Economic benefits of rare and endangered species: summary and meta-analysis', *Ecological Economics* 18, pp. 197-206.

Ricklefs, R. E. y Miller, G. L. (1999), *Ecology*, 4ª ed., New York, Freeman and Company.

Waters, E. C., Holland, D. W. y Weber, B. A. (1994), 'Interregional effects of reduced timber harvests: the impact of the Northern spotted owl listing in rural and urban Oregon', *Journal of Agricultural and Resource Economics* 19(1), pp. 141-160.

Validez interna y externa en la práctica de la Economía Experimental

María Jiménez-Buedo
UNED
mjbuedo@fsof.uned.es

En los últimos años, y sobre todo a partir de los años noventa del pasado siglo, el número de artículos que utilizan el método experimental para la generación de datos primarios y de evidencia empírica en apoyo de tesis teóricas ha aumentado considerablemente en ciencias sociales en las que su uso se creía tradicionalmente difícil o impracticable (como la Economía, la Teoría Sociología y la Ciencia Política) (Levitt y List, 2007; Morton, 2008), e incluso ha surgido una importante rama experimental en filosofía que trata de complementar, y a veces sustituir, a la intuición como fuente primordial en los argumentos acerca de cuestiones morales (Knobe y Nichols, 2008).

Este importante giro en la práctica de las ciencias sociales ha venido acompañado de un esfuerzo paralelo dirigido a reflexionar sobre los aspectos metodológicos y filosóficos que rodean a la práctica experimental en las ciencias sociales. En el caso de la economía experimental esta reflexión metodológica acerca de la práctica experimental ha tenido especial relevancia (Guala, 2005). Esto ha sido así, en parte, por el hecho de que muchos de los resultados de esta disciplina emergente son también relevantes en las discusiones acerca de los supuestos fundamentales sobre la racionalidad de los agentes económicos que cimentan la disciplina económica, y en parte, por tratarse la economía de la disciplina que antes, y con mayor entusiasmo, ha adoptado el experimento como método válido de generación de evidencia empírica, venciendo antes que otras ciencias sociales las tradicionales resistencias a la experimentación en el contexto de lo social.

Una parte fundamental de este esfuerzo de reflexión metodológica acerca de la utilidad y pertinencia de los experimentos en ciencias sociales ha sido, y continúa siéndolo, formulada en torno a las categorías clásicas de validez interna y externa de los experimentos, acuñadas en los trabajos clásicos de Donald T. Campbell y sus colaboradores durante las décadas de los sesenta y los setenta del pasado siglo (Campbell y Stanley, 1963; Cook y Campbell 1979). Siguiendo a Cook y Campbell, la validez interna de un experimento queda normalmente definida como "la validez aproximada con la que podemos inferir que la relación entre dos variables es causal o que la ausencia de relación entre ambas implica la ausencia de causa" (Cook y Campbell, 1979, p. 37) y la validez externa "se refiere a la validez aproximada con la que podemos inferir que una supuesta relación causal puede ser generalizada a otros medidas alternativas de la causa y el efecto y a otro tipo de sujetos, entornos, y momentos" (1979, p. 82; *traducción propia*). En términos más generales, se suele interpretar la validez interna como la propiedad

de un experimento que hace que su diseño justifique extraer la conclusión propuesta acerca de lo que ocurre en dicho experimento, y validez externa como la propiedad de un experimento que hace que su diseño justifique extraer la conclusión propuesta acerca de situaciones o poblaciones concretas fuera o más allá de las condiciones experimentales.

Las categorías cambpellianas de validez interna y externa han sido el centro de un intenso debate por parte de un número no desdeñable de metodólogos asociados a la práctica cuasi-experimental en ciencias sociales. En diálogo con sus críticos, Campbell, junto con varios de sus colaboradores, reformuló paulatinamente la noción de validez experimental y acuñó otras nociones de validez que han sido progresivamente incorporadas a la terminología estándar de la práctica experimental asociada a los cuasi-experimentos de campo (fundamentalmente, las nociones de validez de constructo y de conclusión estadística). El propio Campbell, a partir de la segunda mitad de los años ochenta, y consciente de la controversia y confusión generada en torno a los conceptos de validez interna y externa, declaraba, a propósito del primero de estos términos: "[…]…la validez interna ha acabado siendo asimilada a los experimentos de laboratorio con control total sobre el tratamiento[…]. Dado que esto no es lo que teníamos en mente, tenemos que volver a intentarlo, utilizando nuevos términos" (Campbell 1985, p. 68; *traducción propia*). Para ello, propone un ejercicio de "re-etiquetización" o reclasificación (*relabelling*) de estos términos en los que plantea, tentativamente, sustituir la noción de validez interna por la de validez molar local y la de validez externa por el principio de similitud proximal o de contigüidad (Campbell 1985).

En su gran mayoría, no obstante, los debates filosóficos acerca de la práctica experimental han permanecido prácticamente al margen de estos debates y desarrollos. Asimismo, las nociones de validez interna y externa continúan siendo comúnmente utilizadas por gran parte de los científicos que utilizan el método experimental en sus investigaciones. De esta forma, a pesar de que han sido varios los que han señalado importantes problemas conceptuales alrededor de las nociones de validez interna y externa desde la literatura metodológica asociada al diseño de investigación y experimentación social, estas nociones continúan utilizándose hoy día en la literatura filosófica-científica de forma aproblemática. De hecho, las categorías de validez interna y externa, tal y como fueron formuladas por Campbell, aparecen con frecuencia, explícitamente, en las discusiones acerca de aspectos más problemáticos de la práctica experimental, cumpliendo la función, precisamente, de encarnar una medida de adecuación de los resultados experimentales, si no definitiva, al menos consensual (Caamaño 2009, Cartwright 2006).

La presente comunicación trata de llamar la atención sobre los aspectos más confusos de la díada validez interna-externa y argumenta que la ambigüedad en torno a la interpretación de estos conceptos ha dado lugar a una serie de asociaciones erróneas o equívocos en el contexto de la discusión metodológica acerca de la economía experimental y del comportamiento.

Quizá uno de los aspectos en que más claramente se muestre la falta de claridad conceptual en torno a las nociones de validez interna y externa resida en la confusión que existe alrededor de la caracterización de la relación entre ambas (Jiménez-Buedo, Miller, 2009). De esta forma, es frecuente encontrar en la literatura referencias a dos ideas que parecen en principio incompatibles: de un lado, se defiende que la relación interna es un prerrequisito de la validez externa de un experimento. De otro lado, y más frecuentemente, se defiende la idea de que la validez interna y la externa mantienen una tensión entre sí. Esta idea queda normalmente resumida bajo la noción de un *trade-off* o dilema entre validez interna y validez externa. La idea, intuitivamente atractiva y ampliamente citada en la literatura metodológica acerca de los experimentos en ciencias sociales, fue ya enunciada en los trabajos pioneros de Campbell y sus colaboradores. Así por ejemplo, Campbell y Stanley afirmaron: "ambos tipos de criterio son obviamente importantes, si bien, con frecuencia se encuentran en contradicción, ya que las características [experimentales] que hacen que aumente un tipo de validez, pueden poner en peligro el otro tipo de validez" (Campbell y Stanley 1963). La intuición detrás de este *trade-off* entre validez interna y externa es la siguiente: cuanto más aislemos los factores o variables contaminantes que pueden afectar a la relación causal que queremos identificar, más aseguraremos la validez interna de nuestro diseño, pero a la vez, más se alejará nuestro experimento de las condiciones reales en las que esa relación se produce fuera del contexto experimental, y por tanto, menos generalizables serán los resultados de nuestro experimento a otras situaciones y otras poblaciones. Podemos encontrar numerosas referencias explícitas e implícitas a este *trade-off* o tensión en las discusiones metodológicas acerca del método experimental en economía, psicología social, y las ciencias sociales en general (Brehn et al., 1999; Smith and Mackie, 1999, Guala, 2005).

Esta idea convive, no obstante, y como hemos mencionado, con otra que resulta igualmente atractiva a la intuición, y que si bien, no es tan omnipresente como la noción del *trade-off* entre la validez interna y externa, aparece también con frecuencia en las discusiones metodológicas acerca del método experimental en las ciencias sociales. Se trata de la idea de que la validez interna de un experimento es más bien una precondición de la validez externa y que las cuestiones de validez interna son cronológica y epistemológicamente anteriores a las cuestiones de validez externa (Guala, 2003; Lucas 2003). Hogarth (2005), por ejemplo, sostiene que la validez interna de un experimento es una condición necesaria pero no suficiente para la validez externa, y Thye (2000) ha defendido la idea de que no tiene sentido preguntarse si un resultado se aplica a otros contextos en tanto tengamos dudas acerca de si la relación causal identificada experimentalmente es real o espuria.

Otro problema especialmente importante alrededor de la distinción entre validez interna y externa en el campo de la economía experimental es la falta de acuerdo en la interpretación del concepto de validez externa. En este sentido, en la literatura conviven dos significados claramente distintos de este concepto y por tanto, de qué puede querer decir que un experimento sea externamente válido sólo si sus resultados se dan también *fuera* de las condiciones experimentales. Mientras

que parte de la literatura entiende la validez externa como la seguridad acerca de que los resultados obtenidos en un experimento no son producto de un artificio fruto de las condiciones experimentales, sino que dan cuenta de una relación causal que tiene lugar en otros contexto, otros, como Guala (2005), toman la definición de validez externa como cualidad de un experimento que abre la posibilidad de generalizar a algún sistema concreto o *target system*. Existen sin embargo problemas con ambas definiciones o sentidos de validez externa. En el primer caso, la noción de validez externa sería difícilmente distinguible de la de validez interna. En el segundo caso, sin embargo, el problema residiría en el hecho de que el alcance de una hipótesis causal asociada a un experimento no tiene por qué reflejar su robustez.

Por último, habría que señalar cómo las categorías de validez interna y externa, surgidas en el contexto de cuasi-experimentos de campo asociados en su mayoría a intervenciones ligadas a programas de mejora social, presentan dificultades añadidas cuando son aplicadas a la evaluación de diseños experimentales de laboratorio en el contexto de la economía experimental y del comportamiento. Las definiciones de validez interna y externa de inspiración campbelliana presuponen la existencia de hipótesis causales *asociadas* a la introducción de un tratamiento. Sin embargo, en muchos de los experimentos paradigmáticos en economía del comportamiento como el juego del ultimátum o los juegos de bienes públicos, sin embargo, las hipótesis causales que se proponen para dar cuenta del comportamiento observado, provienen, en cambio, de los factores que el experimentador no controla, fundamentalmente, de la configuración de las preferencias de los sujetos participantes, en lo que Guala, por ejemplo, ha denominado preferencias no-estándar (2008), y que harían referencia a aquellas preferencias que no están conformadas según los supuestos de la teoría de la elección racional clásica. En este sentido, la aplicación de las categorías de validez interna y externa plantea problemas adicionales en el caso de la economía experimental.

Este trabajo concluye abogando por una revisión profunda de la interpretación más común de las nociones de validez interna y externa en la práctica de los experimentos en economía del comportamiento, y por una discusión que traslade aquellos aspectos menos conocidos de los debates metodológicos que han tenido lugar entre los impulsores de la cuasi-experimentación social a la discusión filosófica sobre la práctica experimental.

Referencias bibliográficas

Caamaño Alegre, M. (2009), "Experimental Validity and Pragmatic Modes in Empirical Science", *International Studies in the Philosophy of Science* 23, pp. 19-45.

Cartwright, N. (2006), *Well-Ordered Science: Evidence for Use. Philosophy of Science.*

Campbell, D. T. y Stanley J. C. (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago, Rand McNally and Company.

Cook, T. D. y Campbell, D. T. (1979*), Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Boston, Houghton Mifflin Company.

Guala, F. (2003), "Experimental Localism and External Validity", *Philosophy of Science* 70: 1195-1205.

— (2005), *The methodology of experimental economics*, Cambridge, Cambridge University Press.

— (2008), "Paradigmatic Experiments: The Ultimatum Game from Testing to Measurement Device", *Philosophy of Science* 75, pp. 658-669.

Guala, F. y Mittone L. (2005), "Experiments in economics: External validity and the robustness of phenomena", *Journal of Economic Methodology* 12 (4): 495-515.

Hammersley (1993), "A note on Campbell's distinction between internal and external validity", *Quality & Quantity* 25: 371-387.

— "Abandoning internal and external validity: a response to Swanborn", *Quality & Quantity* 27: 217-218.

Hogarth, R. B. (2005), "The challenge of representativeness design in psychology and economics", *Journal of Economic Methodology* 12 (2): 253-263.

Jiménez-Buedo, M. y Miller, M. (2009), "Experiments in the Social Sciences: The relationship between External and Internal Validity", in [2009 PhilSciArchive] SPSP 2009: Society for Philosophy of Science in Practice (Minnesota, June 18-20, 2009).

Kanazawa, S. (1999), "Using Laboratory Experiments to Test Theories of Corporate Behavior", *Rationality and Society* 11 (4): 443-61.

Knobe, J. y Nichols, S. (2008) *Experimental Philosophy*, Oxford: OUP.

Levitt, S. D. y List, J. A. (2007), "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?", *Journal of Economic Perspectives*, 21 (2): 153-174.

Lucas, J. W. (2003), "Theory-Testing, Generalization and the Problem of External Validity", *Sociological Theory* 21 (3): 236-253.

Morton, R. B. y K. C. Williams. (2008), "Experimentation in Political Science", en Janet Box-Steffensmeier, David Collier y Henry Brady (eds.), *The Oxford Handbook of Political Methodology*, Oxford, OUP.

Schram, A. (2005), "Artificiality: The tension between internal and external validity in economic experiments", *Journal of Economic Methodology* 12 (2): 225-237.

Swanborn, Peter G. (1993), "External validity abandoned?" *Quality & Quantity* 27: 211-215.

Thye, S. R. (2000), "Reliability in Experimental Sociology", *Social Forces* 78 (4): 1277-1309.

Trochim, W. M. K. (1986), *Advances in Quasi-Experimental Design and Analysis*. San Francisco, Jossey-Bass.

Realismo Estructural y carga ontológica de las matemáticas

Carlos M. Madrid Casado
Universidad Complutense de Madrid
carlos.madrid@educa.madrid.org

A medio camino entre el realismo clásico y el instrumentalismo, el realismo estructural ofrece (supuestamente) «lo mejor de ambos mundos», por emplear la expresión de Worrall (1989, p. 99) que ha hecho fortuna. Por un lado, el realista estructural asume que a lo largo de la historia de la ciencia múltiples teorías, que en su momento fueron consideradas como verdaderas a la luz de su éxito, han sido posteriormente refutadas y abandonadas. Por otro, defiende que si las teorías científicas no fueran aproximadamente verdaderas en algún sentido, sería un milagro que realizaran predicciones tan acertadas. La cuestión es cómo acomodar, al tiempo, la «inducción pesimista» y la «inferencia a la mejor explicación». Worrall encontró la solución en Poincaré. A través del cambio científico, el mobiliario que atribuimos al mundo cambia notablemente, pero conserva su disposición. Hay una continuidad en el cambio, pero esa continuidad no es de la ontología, sino de la estructura. Pese a que la ontología individual sufre cambios radicales, las teorías nuevas retienen la estructura matemática de las teorías antiguas, porque las estructuras matemáticas de las teorías científicas representan la estructura del mundo.

El realismo estructural propuesto por Worrall no desborda el ámbito epistemológico, por cuanto sólo sustenta que podemos conocer la forma o estructura del mundo, permaneciendo su contenido o naturaleza incognoscible. Frente a este «realismo estructural epistémico» (REE), el realismo estructural propuesto por French y Ladyman (2003) toma partido por una genuina posición ontológica. El «realismo estructural óntico» (REO) abraza –por decirlo con Psillos (2006, p. 561)– un «estructuralismo óntico» que consiste en creer que «todo lo que *hay* es estructura» y que «nuestras teorías científicas son capaces de capturar las estructuras *existentes* en el mundo».

Ahora bien, en su defensa del realismo científico, tanto REO como REE comparten y se enfrentan a tres problemas graves que requieren su atención:

Problema 1. El realismo estructural da por hecho, al igual que el neopositivismo, que existe una continuidad matemática en la sucesión de teorías científicas (las ecuaciones se preservan o perviven como casos límite), pero la historia de la ciencia aporta numerosos episodios en que la estructura matemática, como la ontología, no ha sobrevivido [cf. Rivadulla (2009) para un análisis exhaustivo de este dogma del realismo estructural].

Problema 2. El realismo estructural da por bueno que la estructura matemática de la teoría es capaz de representar, en sentido realista (no meramente empirista), la estructura del mundo. Pero si las teorías científicas *representan* –por medio de su estructura matemática- la estructura de la realidad, los realistas estructurales tienen qué explicar qué noción de representación emplean para poder garantizar la inferencia realista («las teorías *capturan* la estructura del mundo»). Cuando el realista estructural afirma que la teoría representa el mundo viene a querer decir que la estructura (matemática) de la teoría y la estructura (matematizable) del mundo son (parcialmente) isomorfas. Pero, como he argumentado en Madrid Casado (2008 y 2009), este postulado de isomorfismo estructural entre la teoría y el mundo conduce a un callejón sin salida. Y sin una relación representacional fuerte, los realistas estructurales tienen muy difícil establecer cualquier clase de inferencia sobre la estructura del mundo a partir de la estructura de las teorías.

Problema 3. El realismo estructural acepta, por principio, que podemos distinguir, separar y disociar una estructura y una ontología en nuestras teorías científicas. Sin embargo, la distinción estructura / ontología (forma / contenido, formalismo / interpretación) no es clara ni distinta. No existe un corte limpio entre la estructura y la ontología de nuestras teorías, porque las estructuras matemáticas están cargadas de ontología...

Fue Worrall (1989, p. 117) quien acuñó esta distinción, analizando la evolución de las ecuaciones en óptica: «Fresnel identificó de modo completamente erróneo la *naturaleza* [ontología] de la luz, pero no es ningún milagro que su teoría disfrutase del éxito empírico predictivo que tuvo; no es ningún milagro, porque la teoría de Fresnel, como lo vio la ciencia posterior, atribuyó a la luz la *estructura* correcta». La estructura vendría dada por el formalismo matemático de la teoría; y la ontología, por la interpretación física. Pero, ¿es siempre posible demarcar con precisión la estructura de la ontología?

A mi entender, no existe corte entre ambas y, en consecuencia, no se puede ser realista con respecto a las estructuras y, simultáneamente, antirrealista con respecto a las entidades que contienen. El «semirrealismo», como lo denomina Chakravartty (1998), es altamente inestable; porque uno no puede creer que ciertas relaciones son reales a menos que también acepte que ciertos objetos están relacionados así. La distinción entre la verdad de las relaciones y la verdad acerca de los *relata* no marcha. Por decirlo en unos términos escolásticos que el propio realista estructural resucita: la forma es inseparable de la materia. French y Ladyman (2003, p. 37) responden divorciando el realismo estructural de cualquier compromiso con ontologías de individuos: «una formulación del realismo adecuada a la física precisa estar construida sobre la base de una ontología alternativa que reemplace la noción de objeto por la de estructura». El mundo no consistiría en objetos, sino sólo en estructuras. REO radicaliza –como observa Van Fraassen (2006)- el estructuralismo de REE hasta el punto de «reificar» las estructuras. Pero, ¿es aceptable la desaparición de las entidades, un mundo sin objetos? La indistinguibilidad de las partículas cuánticas mostraría, para sus partidarios, que referirse a ellas es sólo un modo de hablar. Pero los nodos de las

estructuras cuánticas pueden no ser meros fantasmas, sino objetos de una categoría diferente a los macroscópicos. ¿O acaso estos últimos (que podemos señalar y agarrar) tampoco existen, son apariencias? Como concluye Van Fraassen (2007, p. 55): «¿Tiene sentido concebir una estructura que no es estructura de algo? *Una estructura de nada es nada*».

Pero hay más: la ocurrencia del realista estructural de que una misma estructura matemática es compatible con ontologías muy diferentes dista de ser cierta, ya que toda estructura lleva aparejada una *carga ontológica*. La estructura matemática no puede separarse de la ontología física, porque no existe algo así como una neutralidad ontológica de las matemáticas. Ningún lenguaje –ni siquiera el matemático- es neutro en la descripción del mundo.

Por ejemplo: la Mecánica Matricial (MM) de Heisenberg y la Mecánica Ondulatoria (MO) de Schrödinger son dos realizaciones isomorfas de la misma estructura matemática (el espacio de Hilbert), pero las estructuras ónticas que MM y MO prescriben a la realidad no son, ni mucho menos, isomorfas. MM proyecta sobre el mundo una estructura discreta, dado que su realización del espacio de Hilbert –el espacio de las sucesiones de cuadrado sumable– es discreta, lo que llevó a Born y Jordan a referirse a MM como una «verdadera teoría del discontinuo» y acentuaba una visión corpuscular del mundo atómico. En cambio, MO proyecta sobre la realidad una estructura continua, dado que su realización del espacio de Hilbert –el espacio de las funciones de cuadrado integrable– es continua, lo que provocó que Schrödinger hablara de MO como una «nueva física del continuo» y casaba con una concepción ondulatoria del microcosmos. La matemática algebraica de MM apoyaba una ontología discreta, corpuscular. La matemática analítica de MO sugería, por contra, una ontología continua. Si había una ecuación de ondas, tenía que haber ondas o algo similar. Las estructuras matemáticas de MM y MO son equivalentes pero, al ser de muy distinta raigambre, contienen *de facto* subestructuras ónticas incompatibles (discreta *vs.* continua). Confiado en la realidad de la estructura abstracta del espacio de Hilbert, el realista estructural no tiene más opción que ser realista tanto con respecto a la realización discreta como con respecto a la continua; a la manera que el matemático formalista, que trabaja clausurado en su sistema axiomático, no tiene más que aceptar que los números reales poseen tanto modelos estándar –no numerables, continuos- como modelos no estándar –numerables, discretos- (la física reproduce la principal aporía de la matemática: ¿es continuo o discreto el fondo de la naturaleza?). Sin embargo, este paso entraña numerosas dificultades, porque no hay una única estructura óntica ni una única ontología del mundo cuántico asociada con ambos tipos de estructura matemática (dualidad onda-corpúsculo). No basta con aferrarse al formalismo y afirmar que las estructuras matemáticas que éste satisface son las únicas en que cree el realista estructural, porque cada estructura matemática puede conllevar una subestructura óntica distinta. Por encima de que MM y MO nos dibujen un mundo poblado de corpúsculos u ondas respectivamente, ambas teorías ya nos están presentando dos mundos estructuralmente incompatibles (discreto *versus* continuo).

El formalismo condiciona y, a veces, determina la interpretación, porque el aparato matemático empleado arrastra un *peso ontológico*. No quiero decir que la estructura matemática fije de una vez por todas la ontología, pero sí que restringe bastante la clase de ontologías compatibles con el formalismo. La historia de la Mecánica Cuántica nos ofrece más ejemplos de este condicionamiento. En la formulación de la Mecánica Cuántica obra de Feynman, la utilización de la herramienta matemática conocida como integral de camino soporta la interpretación como suma de historias. En la Mecánica Bohmiana, la propiedad de existencia y unicidad global de solución de la ecuación de puntos-guía, que implica que las trayectorias no pueden cortarse ni fusionarse, posibilita la interpretación realista y determinista. O, por no seguir, la interpretación de Born o la de De Broglie, que borran la contradicción entre partículas y ondas, se abrieron paso gracias a la modificación del formalismo (mediante la introducción de medidas de probabilidad y la teoría matemática de la doble solución, respectivamente). Cambian la interpretación y la ontología, porque cambian las estructuras matemáticas. La tarea del físico teórico no se limita a buscar una teoría matemática adecuada a la experiencia, porque el repertorio de teorías matemáticas que tiene a su disposición orienta decisivamente el alcance ontológico.

En resumen, la distinción estructura / ontología colapsa, porque –según la expresión de Psillos– ambas forman un continuo. La *carga ontológica* de las matemáticas no es despreciable y, por consiguiente, la posibilidad de diseccionar la estructura matemática de la ontología de las teorías físicas, como si fueran dos mundos completamente independientes, es un espejismo.

Aparte de dejar constancia de las fisuras del realismo estructural, he tratado de argumentar que las estructuras no pueden separarse de las ontologías, pergeñando una tesis –mucho más próxima a los filósofos continentales que a los analíticos- que atenta directamente contra la viabilidad del realismo estructural: la carga ontológica de la matemática. Una conexión a explorar es la de esta tesis con el argumento de indispensabilidad de Quine-Putnam: en tanto en cuanto las estructuras matemáticas se muestran indispensables para nuestras teorías físicas, pues contribuyen tan genuinamente como lo hacen las partículas hipotéticas, comparten su estatus ontológico. Desde esta perspectiva holista, cuando contrastamos una teoría con la experiencia, no sólo estamos poniendo a prueba las hipótesis físicas, sino también las estructuras matemáticas utilizadas. No en vano, como notara Poincaré, el indeterminismo cuántico rompió el matrimonio entre la física y las ecuaciones diferenciales, al introducir discontinuidades que requerían otra matemática. La matemática puede verse afectada, ontológicamente hablando, por la suerte que corra la física en cuya formulación aparece. Quizá, sencillamente, por la razón de que también aporta su grano de ontología.

Referencias bibliográficas

Chakravartty, A. (1999), ‘Semirealism’, *Stud. His. Phil. Sci.* 29, pp. 391-408.

French, S. y Ladyman, J. (2003), ‘Remodelling Structural Realism: Quantum Physics and the Metaphysics of Structure’, *Synthese* 136, pp. 31-56.

Madrid Casado, C. M. (2008), 'El Realismo Estructural a debate: Matemáticas, Ontología y Representación', *Revista de Filosofía* 33/2, pp. 49-66.

— (2009), 'Do mathematical models represent the World? The case of quantum mathematical models' en J. L. González Recio (ed.), *Philosophical Essays on Physics and Biology*, Hildesheim, Georg Olms Verlag, pp. 67-90.

Psillos, S. (2006), '*The* Structure, the *Whole* Structure, and Nothing *but* the Structure', *Phil. Sci.* 73, pp. 560-570.

Rivadulla, A. (2009), 'Two Dogmas of Structural Realism. A Confirmation of a Philosophical Death Foretold', *sometido*.

Van Fraassen, B. (2006), 'Structure: Its Shadow and Substance', *Brit. J. Phil. Sci.* 57, pp. 275-307.

— (2007), 'Structuralism(s) about Science: Some Common Problems', *Proc. Arist. Soc.* 81, pp. 45-61.

Worrall, J. (1989), 'Structural Realism: The Best of Both Worlds?', *Dialectica* 43, pp. 99-124.

Identity, indiscernibility and naturalised metaphysics

Matteo Morganti
University of Konstanz
Matteo.Morganti@uni-konstanz.de

Introduction

The supporters of the Leibniz-Quine 'reductionist' approach to identity – according to which the identity of objects can be analysed in terms of their properties and ultimately reduces to qualitative uniqueness - have recently insisted on weak discernibility, discernibility determined by irreducible relations, with a view to resisting putative counterexamples to the Identity of the Indiscernibles coming from mathematics and physics. In quantum mechanics, this allegedly makes it possible to resist the widespread view that particles are non-individuals ([Saunders (2006)], [Muller and Saunders (2008)], [Muller and Seevinck (2009)]).

This proposal has been criticised on the basis that it reverses the natural order between relations and relata, but this criticism misses the point: it is not a law of metaphysics that relations depend on their relata. Others [Dieks and Veerstegh (2008)] have argued that the existence of relations can be inferred from 'relation talk' only if specific individuals can be 'picked out' physically without destroying the system (since no such 'symmetry breaking' is possible, these authors conclude that entangled systems have no component parts). But this can consistently be rejected as an unmotivated requirement.

The real worry should be that it is not obvious that the allegedly discerning properties truly are relations. A detailed consideration of the interconnection between quantum theory, its interpretation and metaphysics can shed light on this issue. It also leads to a more general reflection on identity in the context of what is known as 'naturalised' metaphysics.

Quantum properties

In their papers, Muller, Saunders and Seevinck (henceforth, MSS) assume the *countability* of quantum particles: quantum mechanics tell us precisely how many entities one is dealing with when presented with a physical system. On the basis of a minimal set of axioms shared by all interpretations of quantum mechanics, they then argue that identical quantum particles in the same system are 'relationals', i.e., entities that are only weakly discernible. This is because, MSS claim, such particles always partake in genuine, categorical relations, sufficient for discerning them.

MSS maintain that their presented proofs are eminently compelling because the physical significance of the relations they employ is out of question. However, this is not as obvious as they claim.

First of all, consider the formalism used to describe the spin of two identical fermions in the singlet state (a paradigmatic example of the systems examined by MSS). Literally, the 'minimal theory' MSS are eager to limit themselves to says that the whole physical system is in a superposition of two states in each one of which the component fermions have well-defined, opposite spin values; that is, monadic properties. This means that, while one can legitimately say that the two fermions have opposite spin but not a unique, specific spin value, the existence of a genuine, irreducible relation might be questioned.

More generally, MSS appear to make the mistake of embracing naïve realism about operators: starting from projectors corresponding to specific eigenvalues for - admittedly physically meaningful - monadic properties, they construct relations which they then take to be as obviously genuine as the initial monadic properties. But the formal procedure they apply doesn't necessarily preserve ontological genuineness.

But let us grant MSS that they are right a) in taking the countability of quantum particles as uncontroversial; b) in assuming that relations can individuate; and c) in their reconstruction of quantum relations.

Their claims also require d) that these relations be regarded as *strongly non-supervenient*, i.e., actual and yet not depending on/entailing the existence of any property of their relata – as in the case of Lewis' [Lewis (1986)] imaginary 'opposite charge' relation among two particles not possessing separate charges. For, a categorical relation may hold among identical quantum particles in the same system, but the latter certainly do not possess any property grounding such a relation. Muller and Saunders appear to echo an existing claim (see [French (1989)]) to the effect that indeed quantum mechanics exhibits strong non-supervenience when they claim that "[t]he relative direction of components of spin may be well defined, whether or not the directions of those components are themselves defined" [Muller and Saunders (2008), p. 535]. However, one may maintain that, since there are alternative interpretations of the formalism and the MSS reading does *not* follow directly from the theory, one should opt for a different metaphysical interpretation - one not entailing a commitment to such peculiar relations (for instance, a collapse interpretation according to which one has nothing more than a disposition of the total system to exhibit certain properties upon measurement might be considered preferable).

Naturalist metaphysics and the identity of the indiscernibles

A tacit assumption underpinning MSS' arguments seems to be that the weak discernibility of quantum particles is good news because it allows one to stick to a literal reading of the formalism (in particular, of particle labels) while not having recourse to any mysterious metaphysical factor.

In particular, the main (plausible) motivation for insisting on the principle of the Identity of the Indiscernibles (PII) seems to be the 'naturalist' desire not to make any metaphysical assumption not supported by well-corroborated science. In the present case, one would ground claims of identity on scientifically meaningful and empirically accessible qualities of things.

However, naturalism about metaphysics does *not* entail that one must take PII as the criterion of individuation for objects.

First of all, even if one concedes that the naturalist has a reason for thinking that things are exclusively constituted by their properties, PII only follows if properties are additionally conceived of as universals whose instances are numerically identical to each other. This latter claim, however - the basis of realism about universals - is clearly *non-empirical*. The immediate rejoinder that the naturalist must endorse the bundle theory of universals because otherwise s/he has to attribute a mysterious form of non-analysable, primitive identity to property-instances doesn't work. For, certainly, realism about universals immediately accounts for the numerical identity of instances of the same universal; but it doesn't account for the identity of universals themselves. The latter, it seems, must necessarily be considered primitive.

In fact, in view of this it could be suggested that the naturalist can accept primitive identity also for objects. *Contra* Scotus, Ockham famously said that 'haecceitates' (i.e., primitive intrinsic identities) define the 'mode of being' of individual objects without adding anything to them; and this view seems to be what recent defenders of 'primitive thisness' have in mind (see [Adams (1979)]). At least so understood, primitive identity appears acceptable for naturalists, as it just expresses a fundamental fact about a thing, independent of the thing's relationship with other things (in the quantum domain, this fact appears directly mirrored by the countability of quantum particles considered fundamental by MSS!).

One may reply that there are other reasons for insisting on PII. Upon scrutiny, however, it turns out that these are unpersuasive. Consider first other possible a priori reasons. While in Leibniz's philosophy it had to be a priori true that indiscernible distinct individuals cannot exist, we would hardly regard the peculiarly theological arguments for PII formulated by Leibniz as compelling nowadays. And against the claim (put forward in [Della Rocca (2005)]) that if one doesn't assume PII one has to accept implausible scenarios with many identical and co-located objects with all their material parts in common, one can invoke the fundamental mereological law that no two things can share all their parts (statue/piece of bronze distinctions set aside) without equating identity and indiscernibility.

In addition to all this, it is essential to point out that indiscernible objects can make a qualitative, empirical difference, at least as long as their properties are additive: for instance, a world with two identical material objects exhibits twice the mass of a world with only one of them (see [Hawley (2009)]). Hence, it is not the case that the scientifically-minded metaphysician, aiming to only accept

'empirically detectable' metaphysical posits, has any obvious reasons for ruling out indiscernibles.

Haecceitism

As a matter of fact, naturalists themselves acknowledge the existence of valid counterexamples to PII. Ladyman, for instance, considers two-node graphs with no edges and conclude that these mathematical systems contain absolutely indiscernible entities [Ladyman (2007)]. At the same time, however, Ladyman insists that identity may not be grounded on qualitative difference, but must in any case be 'contextual', i.e., extrinsic and determined by the system in which the object possessing it is 'inserted'. The reason for this, he argues, is that otherwise 'haecceitism' (the possibility of entirely non-qualitative differences between distinct worlds) would follow; but haecceitism is contradicted by contemporary science and appears in general in conflict with naturalism (indeed, in the case of the graphs just considered switching the two nodes with each other doesn't give rise to a new mathematical object).

However, primitive identity *need not* entail haecceitistic differences. For, what is true of distinct worlds (roughly, possible configurations for a set of objects) is not univocally determined by the nature of the identity of each individual object inhabiting them. Adams makes this point when he claims that the issue of "whether the identity of the actual philosopher [Aristotle] with the possible tax collector [...] is quite distinct from that of the qualitative or nonqualitative character of Aristotle's identity" [Adams (1979), p. 20].

A counterpart-theoretic treatment of possible worlds, for instance, might be a way of assuming primitive intra-world identities together with anti-haecceitism. Similarly, a Leibnizian may claim that individuals have all their properties essentially and, therefore, are not identical to any other individual in any world.

Additionally, when it comes to accounting for specific facts which apparently contradict the claim that things possess primitive identities, there might be explanations of the evidence alternative to the contextualist view. Consider for example quantum statistics, in which exchanging indistinguishable particles does not give rise to new, statistically relevant states and, therefore, the ontology is deemed significantly non-classical: there, one can claim that particles are more or less traditionally understood individuals (i.e., possess primitive intrinsic identity), but their state-dependent properties (at least in the case of many-particle systems of identical particles) are holistic properties that only belong to the whole and describe correlations (see [Morganti (2009)]).

Even if all this is not deemed convincing, consider the following question: What does it mean exactly for non-qualitatively grounded identity to be contextual? Ladyman suggests that identity and difference be included in the relations characterizing the 'structure' to which objects belong [Ladyman (2007, p. 35)]. But what is the metaphysical counterpart of these relations? In particular, what do these relations reduce to in the case of one-object systems if not to primitive ungrounded identities – albeit of places in structures? It is not clear that

the resulting picture is any less mysterious than the one based on the 'thin', Ockhamist form of primitive intrinsic identity discussed above.

Conclusions

At least on a weak reading of primitive identity, a form of identity not grounded on discernibility might be deemed acceptable even from the naturalist viewpoint. Accepting it allows one to avoid the assumptions required for reaching the conclusion that quantum particles are weakly discernible, at least some of which may legitimately regarded as controversial; and to consider facts of countability sufficient for ascribing individuality.

References

Adams, R.M., (1979), 'Primitive thisness and primitive identity', *Journal of Philosophy* 76, pp. 5-25.

Della Rocca, M., (2005), 'Two spheres, twenty spheres and the Identity of the Indiscernibles', *Pacific Philosophical Quarterly* 86, pp. 480-492.

Dieks, D. and Versteegh, M., (2008), 'Identical quantum particles and weak discernibility', *Foundations of Physics* 38, pp. 923-934.

French, S., (1989), 'Individuality, supervenience and Bell's theorem', *Philosophical Studies* 55, pp. 1-22.

Hawley, K., (2009), 'Identity and indiscernibility', *Mind* 118, pp. 101-119.

Ladyman, J., (2007), 'On the identity and diversity of objects in a structure', *Proceedings of the Aristotelian Society*, Supplementary Volume 81, pp. 23-43.

Lewis, D., (1986), *On the Plurality of Worlds*, Oxford, Blackwell.

Morganti, M., (2009), 'Inherent properties and statistics with individual particles in quantum mechanics', *Studies in the History and Philosophy of Modern Physics* 40, pp. 223-231.

Muller, F.A. and Saunders, S., (2008), 'Discerning fermions', *British Journal for the Philosophy of Science* 59, pp. 499-548.

Muller, F.A. and Seevinck, M., (2009), 'Discerning elementary particles', *Philosophy of Science* 76, in press.

Saunders, S., (2006), 'Are quantum particles objects?', *Analysis* 66, pp. 52-62.

How realist is Structural Realism?

Adam O'Brien
Universitat de València
obrien@alumni.uv.es

One of the main concerns of philosophy of science is how to account for science's empirical success. One way to answer this question would be to simply state that its theories are 'on the right lines' towards representing reality. Another, contrasting view of science's epistemic status, would be that scientific theories are no more than useful fictions enabling us to predict and order the universe but without offering any knowledge as to what 'goes on behind' the phenomena the theories pretend to explain. These two antagonistic positions concerning scientific investigation and theory, namely Realism and Instrumentalism respectively, and the possibility of taking «the best of both worlds», as proposed by John Worrall's aptly named, 'Structural Realism: the best of both worlds' (Worrall 1989 hereafter), will be the main focus of this intervention.

In a later paper Worrall asks «What is it reasonable to believe about our most successful scientific theories such as the general theory of relativity or quantum mechanics? That they are true? Or only that they successfully 'save the phenomena', by being 'empirically adequate'?» (2007a, p. 125) In some sense, of course, the realist will have to defend the 'truth' of such theories. To what extent they are believed to be true or 'approximately' true, is not the question here. The question being: *how* or in *what way* are they true? Worrall continues «In earlier work (Worrall 1989) I explored the attractions of a view called Structural Scientific Realism (hereafter: SSR) This holds that it is reasonable to believe that our successful theories are (approximately) *structurally correct* (and also that this is the *strongest* epistemic claim about them that it is reasonable to make).» (*Ibid*) So here we have Worrall's position clearly stated (1) that the epistemic status of scientific theory is in fact based on the *structure* they represent and (2) that this is the best kind of Realism there is. Or as Psillos has directly criticised «In other words, structural realism, as opposed to scientific realism, somehow restricts the cognitive content of scientific theories to their mathematical structure together with their empirical consequences.» (1995, p. 20) By creating, according to Psillos «a physical and epistemic dichotomy between 'structure' and 'nature'», making only structure knowable, «[SSR] cannot be [considered] the best of both worlds» (*Ibid*).

The problem being that upholding an agnostic position toward the *content* of scientific theories seemingly leaves us with something which is very difficult to understand as any kind of realism at all. How can these structures be defended without also having to defend the objects which they supposedly represent? Or as Psillos strongly states «a likely-to-be-false-best-guess of the furniture of the world

is, if anything, a fig-leaf realism.» (1995, p. 24) Worrall for his part clearly states his realist convictions:

> Whatever esoteric philosophical considerations may be raised, it is difficult to resist the feeling that if a theory can make such a striking, seemingly improbable prediction that nonetheless turns out to be empirically correct, then the theory must somehow be 'approximately true' – it must have somehow latched on, no doubt in an approximate (but nonetheless substantial) way, to the 'deep structure' of the universe: to how things really are in the 'noumenal world' behind or beyond the phenomena. (2007b, p. 4)

Although it is not absolutely clear, as Worrall himself states, that Scientific Realism can «be *inferred* in any interesting sense from science's success» (1989, p. 142), the "no miracles" argument [hereafter: NMA] simply states: it would be a miracle of cosmic coincidence if scientific theories which have managed to predict so much had not also "latched on" to reality, even if it were in some approximate way. What can be said is that NMA challenges the sceptic to explain how else this success is to be accounted for, furthermore, isolated empirical success of one theory is one thing, substantial historical evidence on the other hand, is something else.

This evidence is put into question by such considerations as Kuhn's in *The Structure of Scientific Revolutions*, whereby, as Worrall agrees, «a common-sensical sort of person […] is likely to feel those realist sentiments evaporating if he takes a close look at the *history* of science and particularly at the phenomenon of *scientific revolutions*.» (1989, p. 142) The question, however, could be: does Scientific Realism have to defend itself against paradigmatic shifts in theory?

Since Scientific Realism wishes to defend that scientific theories are empirically successful because they approximate reality in some profound way, when proposed with two empirically successful theories whose assumptions are, nonetheless, logically inconsistent with one another, its judgement as to which of the two theories are closer representations of reality may have to be put on hold. The shift from a Newtonian understanding of the universe to an Einsteinian is one such a case. How, then, are we to account for the empirical success of Newton's theories if they are now understood, on account of their understanding of the forces at play in the universe, to be distinctly un-approximate in comparison with Einstein's?

One such argument could be to appeal to Newton's theories as limiting cases of Einstein's; perhaps on some less galactic, terrestrial level. This, however, seems to be missing the point. Newton's "action-at-a-distance" notion of gravity and Einstein's "curvature in space-time" are, at the least, very far removed from one another and consistency between them, therefore, difficult to reconcile. If, according to the pessimistic meta-induction consideration [hereafter: PMI], all past theories are shown to be logically inconsistent with current science, as could be argued from the existence of paradigmatic change, then who is to say that our current theories will not also suffer the same fate? The point being that perhaps the

empirical success of scientific theory, current theory being included, after considering the case for scientific revolutions ought to be understood in miraculous terms after all. This would 'leave the door open' as it were for instrumental or pragmatic positions to stake their anti-realist claims. As Worrall concludes:

> The pessimistic meta-induction, if accepted, would trump the NMA. This is because the history of science would then, as already noted, provide a list of alleged 'miracles' […] This is why it is important for a defensible realism to establish a way in which successive theories in mature science have indeed been at least quasi-accumulative. And I claim that only the structuralist can successfully establish such an account. (2007a, p. 147)

Furthermore:

> …what surely needs to be constructed in response to the 'pessimistic' challenge is a middle position that identifies a level, above that of the observational generalisations, at which scientific progress has been cumulative (or 'essentially' cumulative), despite 'scientific revolutions': and goes on to assert an unambiguous 'realism' about science at the level thus identified. (2003, p. 235)

Worrall believes to have found this example *par excellence* in the shift from Fresnel's "lumiferous aether" to Maxwell's electro-magnetic field, wherein the mathematical equations in question, representing relationships expressing light refraction, were kept entirely intact despite the drastic change on the purely theoretical level. Although this may be the case, the question of whether this should be considered Scientific Realism at all still remains. Worrall, nevertheless, quotes Poincaré on this subject of theory shifts between seemingly incommensurable subsequent fields of investigation which maintain, nonetheless, certain specific structural similarities:

> They teach us now [mathematical equations carried over to subsequent theories], as they did then, that there is such and such a relation between this thing and that; only the something which we then called *motion*, we now call *electric current*. But these are merely names of the images we substituted for the real objects of Nature which will hide for ever from our eyes. The true relations between these real objects are the only reality we can attain. [(Poincaré 1905, p.162) 1989, p. 158].

Claiming that our theories "are merely names of the images we substituted for the real objects of Nature which will hide for ever from our eyes", does not seem, however, to improve Worrall's realist cause. It does in fact have much more in common with Van-Frassen's brand of agnostic instrumentalism [van Fraassen, 1980] and seems to tie in with Psillos' aforementioned criticism that SSR in fact «restricts the cognitive content of scientific theories». Parallels could also be made to a Kantian 'limiting of human knowledge' with Worrall's reference to a «noumenal» world; whereby scientific knowledge is reduced to a structural level only revealed to us by theories which are in themselves inadequate representations of this ultimately unknowable reality – we do not know what space or time is; although we know them to exist – in a similar way:

> From the vantage-point of Maxwell's theory as eventually accepted, this account [Fresnel's], to repeat, is entirely wrong. How could it be anything else when there is no elastic ether to do any vibrating? None the less, from this vantage-point, Fresnel's theory has exactly the right structure – it's 'just' that what vibrates according to Maxwell's theory are the electric and magnetic field strengths.[...] None the less, Fresnel was quite right not just about a whole range of optical phenomena [the famous white spot for example], but right that these phenomena depend *on something or other* that undergoes periodic changes at right angles to light. (1989, p.159) [the italics are mine]

Apart from the fact that this "something or other" does not seem to be much of a realist claim. The question remains: how are we to be sure that this "something or other" has in fact latched onto the deep structure of the universe? So much so, in fact, for Worrall to claim «The structural realist simply, asserts, in other words, that, in view of the theory's enormous empirical success, the structure of the universe is (probably) something like quantum-mechanical.» (1989, p. 163) Although, as Worrall has admitted, the Fresnel-Maxwell shift in optic theory is «unrepresentative» and that «the mathematics of any theory replaced in a 'scientific revolution', while not being retained fully intact, is instead 'quasi-retained' *modulo* the 'correspondence principle'» (2007a, p. 142); «In virtue of what [Psillos insists] do Maxwell's equations correctly represent the structure of the field?» (1996, p. 25) Or, as we could also ask: why should we consider, as Worrall claims, the relationships between phenomena expressed in Newton's theories to be «genuine primitives»? (1989, p. 162)

> According to the account of theory change that underpins SSR, successive theories in science have not only been successively more empirically adequate, but there has always been a *reason*, when viewed from the vantage point of the later theory, *why* the earlier theory achieved the degree of empirical adequacy that it did namely that the earlier theory continues to look approximately correct: its mathematical equations are retained *modulo* the correspondence theory. (2007a, p. 143)

Is this any explanation at all? Should we now be on the look out for other relationships expressed by mathematical equations which have also survived drastic revolutionary shifts in theory? Does this structural 'carry over' actually give more potential to a theory's being accepted as approximately correct? Worrall has warned against «wholesale» statements of science, instead vouching for individual «'retail' arguments for realism about particular theories that have established the 'maturity' of their field by proving predictably successful» (2007b, p. 8), such as the Fresnel-Maxwell shift and subsequent quantum developments in optics theory which then lead on to further development of Quantum Mechanics. Although SSR takes up the challenge of PMI, showing how theoretical continuity may be possible, without a description of what exactly a structure may actually be like and how a scientific theory may latch onto it, Worrall's Realism case seems weak. In as much as suggesting an interesting new approach concerning scientific empirical success and how any Realist position must take up its defensive stance,

Worrall's position is definitely interesting and could well prove to be a promising subject of investigation.

References

Kuhn T S. (1996) *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press.

Psillos S. (1995) 'Is Structural Realism the Best of Both Worlds', *Dialectica*, 49: 15-46.

van Frassen B. (1980) *The Scientific Image*, Oxford, Clarendon Press.

Worrall J. (1989) 'Structural Realism: The Best Of Both Worlds?', *Dialectica*, 43: 99-124.

Worrall J. (2003) 'Tracking Track Records II: Relying on meta-induction', *The Aristotelian Society Supplementary Volume*, 74 (1): 207-235.

Worrall J. (2007a) 'Miracles and Models: Why reports of the death of Structural Realism may be exaggerated', *The Journal of the Royal Institute of Philosophy*, 61 (Supp): 125-154.

Worrall J. (2007b) 'Miracles, Pessimism and Scientific Realism', *British Journal for the Philosophy of Science:* 1-55.

Tiempo, física y libre albedrío*

Daniel Quesada
Universitat Autònoma de Barcelona
daniel.quesada@uab.cat

En Hoefer (2002) se defiende una posición compatibilista entre la física y el libre albedrío. Sostiene Hoefer que "[el] desafío al libre albedrío a partir del determinismo no ha procedido de la física, sino más bien de un matrimonio *non-sancto* de la física con nuestra concepción del tiempo, que incorpora la serie-A" (p. 206). Simpatizo con esta posición compatibilista, pero creo que presenta una dificultad muy importante que tiene que ver con una incompatibilidad subyacente, la que presuntamente se da entre nuestra concepción (corriente) del tiempo y la concepción de la física. Este trabajo pretende exponer la relevancia de esta otra incompatibilidad para una posición compatibilista sobre el libre albedrío como la de Hoefer y explorar brevemente las posibilidades de superarla.

I. La (presunta) incompatibilidad entre nuestra concepción común del tiempo y la de la física descansa en la incompatibilidad existente entre dos concepciones del tiempo. Según una de ellas, los momentos de tiempo están estructurados simplemente como un orden parcial. Según la otra, el tiempo es esencialmente algo que *pasa*. Un acaecimiento está primero en el futuro, luego se hace presente y finalmente se sitúa en el pasado. Es usual sostener que estas concepciones del tiempo corresponden a tipos diferentes de tiempo: respectivamente, el tiempo de la serie-B, y el tiempo de la serie-A, en la denominación consagrada desde el análisis que McTaggart realizó en *The Nature of Existence*, donde se describe claramente la distinción.

Como suele afirmarse, el tiempo de nuestra experiencia común y corriente —el tiempo en que "vivimos nuestras vidas"— es el tiempo de la serie-A, o, más claramente, el tiempo de la serie-A es el tiempo de nuestra concepción común. Científicos famosos, como Einstein o Gödel, así como la mayoría de los filósofos de la ciencia, piensan que no hay nada así en la noción del tiempo de la física contemporánea. A partir de aquí un puñado de filósofos concluyen que la física es incompleta o incluso que presenta fallos, pero la inmensa mayoría extraen la conclusión opuesta, a saber, que el tiempo de nuestra concepción común es una ilusión. Que realmente no existe nada así.

En el trabajo mencionado, Hoefer sostiene que una vez que, finalmente, se produce el divorcio entre la física y nuestra concepción (común) del tiempo

* Este trabajo se ha beneficiado de la financiación del Ministerio de Ciencia e Innovación a través del proyecto de investigación FFI2008-06164-C02-02 y de la ayuda de la Generalitat de Catalunya al Grup d'Investigació en Epistemologia i Ciències Cognitives (GRECC) SGR2009-1528.

"estamos perfectamente justificados en ver nuestras propias acciones *no* como determinadas por el pasado, *ni tampoco* como determinadas por el futuro, sino como, simplemente, determinadas (en la medida en que es adecuado aplicar este término) *por nosotros mismos, por nuestra propia voluntad*" (*op. cit.*, p. 208; el énfasis es suyo). Pero, como he mencionado ya, mantiene igualmente que "la serie-A domina… nuestro pensamiento" (*id.*, p. 208). Si esto es realmente así, entonces difícilmente puede escapar a este dominio el modo en que concebimos el libre albedrío, puesto que éste tiene aspectos temporales esenciales. Es decir, la determinación de la que habla Hoefer puede darse de este modo: podemos "determinar" *ahora* llevar a cabo un cierta acción en un determinado *futuro*. Si se piensa, como Hoefer mismo piensa, que la física es incompatible con la concepción del tiempo de la serie-A y que ante un conflicto entre la concepción de la física y la de nuestro modo común de concebir las cosas es la primera la que debe prevalecer, nuestra concepción del libre albedrío —precisamente por llevar implícitamente incorporada nuestra concepción común del tiempo, presuntamente incompatible con la física— se ve, después de todo, desacreditada.

Veamos de modo pormenorizado cuáles son los supuestos que llevarían a este descrédito. Son, según se ha expuesto, los siguientes:

(i) En caso de conflicto entre la concepción del tiempo de la física contemporánea y nuestra concepción común del tiempo la primera debe prevalecer.

(ii) La física contemporánea es incompatible con el tiempo de la serie-A.

(iii) Nuestra concepción del libre albedrío tiene aspectos temporales esenciales y estos aspectos reflejan nuestra concepción común del tiempo.

(iv) Nuestra concepción común del tiempo es la del tiempo de la serie-A.

II. ¿Cuál son las perspectivas para la superación de esta situación, después de todo desfavorable para una posición compatibilista sobre el libre albedrío?

Alguien podría sostener que (i) es un caso descarado de cientifismo. Realmente esto es lo que parecen pensar los filósofos que, aceptando (ii) y (iii), sostienen que es la *física contemporánea* la que debe modificarse. Pero esta posición extrema parece retrotaernos a una época en la que se consideraba a la metafísica como la reina de las ciencias, aunque sea aquí bajo la guisa de metafísica descriptiva de nuestro sistema conceptual común.

Más claramente incuestionable aún parece el supuesto (iii). Es obvio que nuestra concepción de una acción libremente realizada, como la concepción general de una acción, presenta aspectos temporales, y no se ve desde dónde podría proponerse que éstos constituyen una excepción a nuestra concepción común del tiempo.

Mayor atractivo ofrece, en principio, la posibilidad de cuestionar el supuesto (ii). Y, en efecto, recientemente algunos filósofos de la física han tratado de caracterizar nociones de paso del tiempo o de transcurso o fluir del tiempo dentro de la física relativista.

Así, por ejemplo, Tim Maudlin ha argumentado que hay una asimetría intrínseca en la estructura del espacio-tiempo, y que el paso del tiempo debería identificarse con esta asimetría [cf. Maudlin 2007, capítulo IV]. Un camino diferente es el que sigue Steven Savitt al caracterizar el transcurrir del tiempo como el "sucesivo darse de 'ahoras' (*nows*) locales a lo largo de una curva temporal", donde estos "ahoras locales" se defien con la ayuda de la noción de cono de luz y el supuesto de que el tiempo tiene una dirección, que permite formular las nociones de conos de luz *futuros* y *pasados* [cf. Savitt 2008, § V]. Otros filósofos, como Abner Shimony (1993) y Dennis Dieks (2006), han presentado propuestas que están conceptualmente relacionadas con la de Savitt, si bien son incompatibles con ella.

Los propios proponentes de estos diversos modos de hacer compatibles la física relativista con la idea de paso del tiempo reconocen que se enfrentan a problemas importantes, o, al menos, que sus propuestas se encuentran en una fase inicial y tentativa. Véase, por ejemplo, el siguiente texto revelador de Savitt: "Estoy… suponiendo que una de las orientaciones se ha elegido como el futuro (y la otra como el pasado). No sé cómo se selecciona esta orientación. Quizá se base la elección en alguna asimetría entre las leyes fundamentales de la física, pero se trata de un problema (profundo) a tratar en otra ocasión. Supondré que tenemos dada una orientación" (*loc. cit.*). En Maudlin (2007) se selecciona una orientación sobre la base de la noción de paso o transcurso (del tiempo), pero esta noción misma se toma como primitiva.

Está claro que no podemos conformarnos con las consideraciones —positivas o negativas— que los propios autores hagan acerca de sus propuestas, y que la cuestión debe estar sujeta a un análisis independiente. A mi parecer hay al menos una objeción fundamental a la aceptación en la física de la idea de que el tiempo pasa o fluye a la que no se han enfrentado aún de forma satisfactoria las propuestas en cuestión, y es la objeción de que, si aceptáramos que el tiempo fluye, tendría pleno sentido preguntar con qué velocidad lo hace [cf. Price 1996, p. 13], y, realmente, esto es algo que no parece tener ningún sentido objetivo. Con todo, probablemente aún es pronto para pronunciarse de un modo taxativo sobre las mencionadas propuestas.

La única posibilidad restante consistiría en poner en cuestión el supuesto (iv). A primera vista, este cuestionamiento parece, ciertamente, descabellado. ¿Hay, podría preguntarse, algo más central a nuestro modo usual de pensar sobre el tiempo que lo que se expresa en la expresión común "el tiempo pasa"? Lo que estas expresión, y muchas otras como ella revelan es que es indudable que, en un cierto sentido, expresamos en el lenguaje un compromiso con una concepción del tiempo de la serie-A. Sin embargo, quizá podría afirmarse que tales expresiones no deben tomarse en un sentido literal, por lo que no son definitivamente reveladoras de un compromiso último de nuestra concepción común del tiempo. ¿A dónde acudir entonces para averiguar si esta concepción del tiempo está, realmente, comprometida con la de la serie-A? Sorprendentemente, nuestra *experiencia* del tiempo se ha ofrecido como una alternativa. En efecto, en la vía

quasi-compatibilista —completamente distinta a la mencionada anteriormente— que se sigue en los enfoques recientes de Butterfield (1984), Callendar (2008), y Huggett (en preparación) se pretende reconciliar la "imagen manifiesta" con la "imagen científica" del tiempo sosteniendo que nuestra experiencia temporal no es incompatible con el tiempo de la teoría física. En particular, el análisis de experiencias temporales como la experiencia del "ahora" por Butterfield y Callendar, y la experiencia del "dinamismo" —especialmente del movimiento— por Huggett apunta a que dichas experiencias no nos comprometen con la idea de paso de tiempo.

En mi opinión sería conveniente proseguir esta vía compatibilista sosteniendo que la *noción* común del tiempo no está realmente comprometida con la idea de paso (tiempo de la serie-A). En relación con las experiencias temporales examinadas por los autores mencionados, se podría argumentar entonces que tales experiencias, o bien no son realmente constitutivas de la noción común de tiempo o bien, si lo son, no comprometen a esa noción con la idea de paso de tiempo (los propios argumentos de los filósofos aludidos pueden utilizarse para este último fin).

Una clave para examinar la cuestión de la compatibilidad la suministraría el examen de la experiencia perceptiva de objetos temporalmente extensos, como la experiencia perceptiva de un sonido o de una melodía. La sugerencia es que el análisis de este tipo de experiencia revelaría con claridad que lo que concebimos como cambiante es únicamente nuestra propia perspectiva temporal.

Pues bien, aunque la noción común del tiempo no sea un concepto perceptivo, sí puede estar informada por estos cambios de experiencia temporal (quizá junto con otras experiencias como las anteriormente mencionadas). Y si nuestra concepción común del tiempo está informada por estos cambios de experiencia temporal, es decir, si comprender nuestra noción común del tiempo exige tener la experiencia de ese cambio de perspectiva, entonces, no estaría (literalmente) comprometida con la idea del paso del tiempo. De este modo podría, después de todo, abrirse una (nueva) vía para resistirse a la conclusión de que la física actual desacredita la idea de libre albedrío por su (presunta) incompatibilidad con los aspectos temporales implícitos en esta idea.

Referencias bibliográficas

Butterfield, J. (1984), 'Seeing the Present', *Mind* 93, pp. 161-176.

Callender, C. (2008), 'The Common Now', *Philosophical Issues* 18, pp. 339-361.

Dieks, D. (2006), 'Becoming, Relativity and Locality', en Dieks, D. (ed.), *The Ontology of Spacetime*, Volume 1, Amsterdam, Elsevier.

Hoefer, C. (2002), 'Freedom from the Inside Out', en Callender, C. (ed.), *Time, Reality and Experience*, Royal Institute of Philosophy Supplement 50, Cambridge, Cambridge University Press.

Huggett, N. (en preparación), *Everywhere and Everywhen: Adventures in Physics and Philosophy*.

Maudlin, T. (2007), *The Metaphysics Within Physics*, Oxford, Oxford University Press.

McTaggart, J. (1927), *The Nature of Existence*, Volume II, Cambridge, Cambridge University Press. Extractos relevantes del capítulo 33 de este libro son accesibles con el título de 'The Unreality of Time', en Le Poidevin, R. y MacBeath, M. (eds.), *The Philosophy of Time*, Oxford, Oxford University Press, 1993.

Price, H. (1996), *Time's Arrow and Archimedes' Point*, Oxford, Oxford University Press.

Savitt, S. (2008), 'The Transient *nows*', en Myrvoid, W. C. y Christian, J. (eds), *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony*, Berlin, Springer.

Shimony, A. (1993), 'Reality, Causality, and Closing the Circle', en Shimony, A., *Search for a Naturalistic World View,* Volume I, Cambridge, Cambridge University Press.

El papel de las combinaciones conceptuales en los diseños experimentales

Iván Redondo Orta
Universitat Autònoma de Barcelona
Ivan.Redondo@campus.uab.cat

El desarrollo de la ponencia estará dividido en tres partes que constituirán su base argumentativa. En primer lugar, se partirá del enfoque de la naturalización de la filosofía de la ciencia. La filosofía de la ciencia del Círculo de Viena estaba basada en el análisis lógico-formal o sintáctico de las teorías científicas y consideraba que su tarea consistía exclusivamente en el estudio y elaboración de juicios analíticos, disociados de los sintéticos o empíricos. De este modo, la filosofía se situaba en un lugar privilegiado desde el que analizar y valorar dichas teorías. Quine puso en tela de juicio esta radical disociación entre juicios analíticos y sintéticos, y condenaba así a la filosofía de la ciencia a una circularidad, ya que sus análisis englobaban también aspectos relacionados con el mundo empírico.

La naturalización de la filosofía de la ciencia consiste en aceptar esta tesis y recurrir a las ciencias que estudian al ser humano utilizando sus resultados empíricos para hacer filosofía de la ciencia. Quine utilizó la psicología conductista, mientras que Kuhn se basó en la Gestalt. En nuestro caso, las ciencias cognitivas, debido a su gran desarrollo en las últimas décadas, serán la base científica para nuestra filosofía de la ciencia naturalizada (Giere o Nersessian también hacen una filosofía de la ciencia naturalizada, basada en esta área).

En segundo lugar, dentro de la filosofía de la ciencia naturalizada, muchos autores, como Kuhn o Feyerabend, son susceptibles de interpretaciones constructivistas. Estos autores comparten, junto con el Círculo de Viena, un enfoque teoreticista. Nuestro análisis estará en consonancia con la "tradición experimental en filosofía de la ciencia", una tendencia emergente en los últimos años que, centrándose en el papel de los experimentos y su importancia, se aleja de la tendencia teoreticista y constructivista anterior.

Finalmente, y como punto de controversia al finalizar la charla, se planteará el problema de la normatividad. Dentro de una filosofía de la ciencia puramente analítica, es fácil conceder a esta disciplina una tarea normativa, en tanto en cuanto detecta errores lógico-formales en las teorías científicas, independientemente de los contenidos. Sin embargo, dentro de una filosofía de la ciencia naturalizada, en la que se recurre a resultados de unas ciencias empíricas para analizar problemas de las mismas u otras ciencias empíricas, la circularidad argumentativa choca con el rol normativo de la filosofía de la ciencia. Al final de la conferencia plantearemos una posible salida a esta aparente contradicción.

Con respecto al primer punto, en ciencias cognitivas se puede hablar de dos estilos o tendencias, a saber, el cognitivista y el conexionista (o en red). El primero

tiene la computadora como modelo y se basa en el procesamiento en base a reglas formales de símbolos que designan entidades. Por el contrario, el enfoque conexionista tiene como modelo el cerebro o sistema nervioso del organismo y basa la cognición en propiedades emergentes de la configuración e interacción de todos los componentes del sistema. En la conferencia se explicarán las diferencias e implicaciones con algo más de detalle. Aquí, a modo de resumen, cabe destacar que el primero se basa en relaciones de identidad, y el segundo en relaciones de semejanza o familiaridad. Antes de la identidad entre símbolo y entidad, hay una integración y una generación de sentido por parte del cerebro. Este proceso de integración de información se lleva a cabo de una manera inconsciente o preconsciente. El modelo conexionista ha tenido como base empírica muchos estudios neurocientíficos en los últimos tiempos.

Dentro de este estilo conexionista o en red, la teoría de los "*conceptual blendings*" (combinaciones conceptuales) de Fauconnier y Turner, junto con la noción de "*material anchors*" (anclajes materiales) para estas combinaciones conceptuales, de Hutchins, serán los pilares básicos que utilizaremos para nuestro análisis de metodología científica. La teoría de las combinaciones conceptuales sostiene que los seres humanos generamos esquemas o patrones neuronales (*schemas*, *frames*) de diferentes situaciones o escenarios de la vida cotidiana. El cerebro genera a su vez relaciones de semejanza, no de identidad, entre estos diferentes escenarios, lo que permite crear espacios mentales genéricos con elementos que esos diferentes espacios mentales guardan en común. Finalmente, del espacio genérico emerge un espacio mental donde se combinan de manera selectiva elementos de ambos espacios, y se da lugar a un espacio con propiedades emergentes propias.

El ejemplo del debate con Kant, de los mismos autores, servirá como aclaración de la teoría. Imaginemos un filósofo actual que, mediante resultados empíricos de ciencias cognitivas, menciona tesis de Kant para contraponerlas a dichos resultados. En este espacio, Kant está muerto, el profesor vivo, hablando en inglés, utilizando procesos cognitivos para hallar la verdad sobre algunas cuestiones, etc. El otro espacio mental en cuestión es el de la época moderna, en el que Kant, en alemán, buscaba la verdad a través de la razón, y el profesor actual, así como las ciencias cognitivas, no existían como tales. En el espacio genérico, mediante relaciones de semejanza, tenemos dos profesores, dos lenguajes, dos tiempos y espacios y la búsqueda de la verdad. En el espacio emergente de selección proyectiva, tenemos la situación actual, con dos profesores debatiendo con argumentos y contraargumentos, en inglés, y buscando la verdad en base a diferentes métodos o aproximaciones que chocan entre sí. Este espacio combinado tiene propiedades emergentes que lo convierten en un debate elaborado entre dos profesores que de otro modo jamás hubieran podido entrar en contacto.

Fauconnier y Turner sostienen que las combinaciones conceptuales descritas tienen la función de comprimir relaciones vitales (que son cambio, identidad, tiempo, espacio, causa-efecto, parte-todo, representación, rol, analogía, disanalogía, propiedad, semejanza, categoría, intencionalidad y unicidad) para generar espacios a escala humana, que permiten la comprensión por parte del ser

humano de procesos o fenómenos que escaparían a la misma. La imaginación está en la base de esta compresión de relaciones vitales. En el ejemplo del debate con Kant, la semejanza entre lenguajes (alemán e inglés) se comprime en identidad (inglés), se genera una relación de intencionalidad entre ambos profesores, se comprimen tiempo (el presente) y espacio (el aula del profesor), etc. La compresión de todas estas relaciones vitales permite generar un debate imaginario accesible a las capacidades cognitivas humanas y a su capacidad de transmisión y comunicación.

Dado que estas combinaciones conceptuales son fruto de la imaginación y de la capacidad integradora y generadora de sentido del cerebro, podrían considerarse espacios mentales emergentes efímeros. No obstante, algunas de estas combinaciones conceptuales se convierten en intersubjetivas y persisten a lo largo del tiempo en sociedades o culturas de una forma estable. Esto es gracias a lo que Hutchins denomina anclajes materiales. La obra de un artista es un anclaje material de sus combinaciones conceptuales, por ejemplo. Hutchins, Fauconnier y Turner analizan anclajes materiales como el reloj, los billetes y monedas, los dedos como "calendario" en el caso de los japoneses, etc., así como sus combinaciones conceptuales subyacentes. Que una combinación conceptual tenga elementos materiales integrantes es vital para su persistencia y robustez.

Con todas estas aclaraciones, analizaremos el segundo punto a tratar. La exposición estará en consonancia con la "tradición experimental" en filosofía de la ciencia, debido a su crítica e intento de superación del constructivismo o subjetivismo. No obstante, esta tradición, que ha estudiado fundamentalmente las ciencias "duras" (física, química, biología) de la mano de autores como Hacking, Rheinberger, etc., se basa en el papel de los artefactos y otros elementos materiales que entran en juego en los experimentos científicos. Sin embargo, dejan de lado o en segundo plano (quizás por contrarrestar el constructivismo mencionado) el papel de la imaginación o los procesos cognitivos de los científicos implicados en tareas experimentales (a excepción de Gooding, que analiza el caso de Faraday atendiendo a aspectos tanto materiales como perceptivos e imaginativos). Recientemente, en el libro editado por Radder, hay autores que reclaman este análisis combinado. La teoría de las combinaciones conceptuales (aspecto mental/imaginativo de la cognición), junto con los anclajes materiales (aspecto material), se mostrará como el modelo más eficaz para analizar los diseños experimentales combinando aspectos internos o mentales, y externos o relacionados con el diseño y manipulación del entorno. Durante la ponencia, se utilizarán como casos o ejemplos diseños experimentales de psicología.

Los diseños experimentales en psicología, en esencia, utilizan diferentes sujetos experimentales (diseño intersujeto) o un mismo sujeto en diferentes fases (diseño intrasujeto). En ambos casos, el diseño experimental intenta captar relaciones causales entre una variable independiente y una variable dependiente. En el diseño intersujeto, por ejemplo, cada sujeto experimental, con sus características y cualidades, constituye un espacio. En el espacio mental genérico, todos son sujetos con cualidades más o menos semejantes entre sí. En el espacio emergente que constituye el diseño experimental, sin embargo, todas estas

cualidades se convierten en variables con diferentes valores posibles. Las variables que no interesa que intervengan en los resultados se convierten, en el espacio mental combinado, en variables enmascaradas, que deben ser controladas. Veremos algunas estrategias de control de estas variables, como la manipulación o la aleatorización, para asegurar la legitimidad de las combinaciones llevadas a cabo en el diseño, ya que una ausencia de control de estas variables puede dar lugar a una distorsión en la interpretación de los resultados. Tal y como sostiene la teoría de las combinaciones conceptuales, veremos ejemplos de diseño experimental en los que se comprimen relaciones vitales como las de espacio, tiempo, las de cambio o semejanza (comprimidas en identidad), causalidad, etc. Todas estas compresiones se llevan a cabo a través de controles experimentales, que interpretaremos como anclajes materiales para estas combinaciones conceptuales. Además, veremos como la operativización de las variables también consiste en una combinación conceptual que relaciona acciones concretas y medibles de esos sujetos con procesos internos, conceptuales o abstractos. Esas acciones concretas en las que se traducen los procesos mentales serán consideradas como anclajes materiales.

Tras este análisis, la ponencia finalizará con la tercera parte que hace referencia al problema de la normatividad. La filosofía naturalizada basada en las combinaciones conceptuales tiene un componente normativo intrínseco, ya que la filosofía puede y debe ejercer una labor crítica con las combinaciones conceptuales subyacentes a los diseños experimentales. De hecho, muchos artículos de los propios psicólogos están destinados a desenmascarar las variables enmascaradas tras estas combinaciones conceptuales. La noción de "Deblend" ("descompresión") de Fauconnier y Turner, utilizada en su obra, será de ayuda para analizar esta labor normativa de la filosofía de la ciencia.

Referencias bibliográficas

Estany, A. (2001), *La Fascinación por el Saber. Introducción a la Teoría del Conocimiento*, Barcelona, Crítica.

Fauconnier, G. y Turner, M. (2002), *The Way We Think: Conceptual Blending and the hidden Complexities in Mind*, New York, Basic Books.

Hutchins, E. (2005), 'Material Anchors for Conceptual Blendings', *Journal of Pragmatics* 37: 1555-1577.

León, O. G. y Montero, I. (2003), *Métodos de investigación en psicología y educación*, Madrid, McGraw Hill.

Radder, H. (2003), *The Philosophy of Scientific Experimentation*, Pittsburgh, University of Pittsburgh Press.

La *preducción teórica*, una práctica deductiva de descubrimiento científico

Andrés Rivadulla[*]
Universidad Complutense de Madrid
arivadulla@filos.ucm.es

Introducción

Durante el siglo XX los filósofos de la ciencia han prestado muy poca atención al contexto de descubrimiento. Desde la consolidación del método hipotético-deductivo por Einstein y Popper, y tras la reincorporación de la inducción al contexto de justificación, *via* la lógica inductiva de Carnap, los metodólogos de la ciencia han descuidado en gran medida el hecho de la creatividad científica. La condena de Popper de las formas en que nuevas ideas acceden a la ciencia, y su solución negativa del problema lógico y metodológico de la inducción, contribuyeron a su abandono.

Por esta razón no resulta extraño que muchos teóricos de la ciencia ignorasen también a la abducción, una inferencia igualmente ampliativa y estrechamente ligada a la inducción, y que ha sido una práctica de descubrimiento usada en las ciencias observacionales de la naturaleza y en las partes empíricas o fenomenológicas de las ciencias teóricas.

Este abandono no impidió empero que en los años ochenta del siglo pasado se produjese una renovación del interés por la abducción Peirceana, y la cuestión de la existencia de una lógica del descubrimiento atrajo la atención de científicos, si bien en dominios como la inteligencia artificial, la adquisición de conocimiento, la programación lógica y materias afines. Por otra parte, desde la identificación por Harman de la abducción con la inferencia a la mejor explicación, la abducción ha sido considerada también como una parte fundamental del argumento a favor del realismo científico.

Subrayo la importancia de la abducción Peirceana para el contexto de descubrimiento en la metodología de las ciencias observacionales de la naturaleza. Pero discrepo de Peirce respecto de que la abducción sea la única forma de razonamiento por medio de la cual ideas nuevas acceden a la ciencia. Y en particular rechazo su idea de que la deducción no puede producir nunca ideas nuevas. Antes al contrario, asevero que el razonamiento deductivo puede ser aplicado al contexto de descubrimiento en la metodología de las ciencias teóricas, como la física matemática. De hecho, Peter Medawar, premio Nobel de Medicina en 1969, reconoció en su contribución al volumen de Schilpp dedicado a Karl

[*] Grupo de Investigación Complutense de Filosofía del Lenguaje, de la Naturaleza y de la Ciencia, Ref.930174-603.

Popper, que la renuncia a explicar cómo se originan las hipótesis constituía un signo de la debilidad del deductivismo hipotético. Así, denomino *preducción* a la forma de razonamiento consistente en la implementación del razonamiento deductivo al contexto de descubrimiento científico. El objetivo de mi contribución precisamente es mostrar el papel que la preducción juega para la creatividad en las ciencias físicas.

Afirmo que la *preducción* es la forma de razonamiento que, partiendo de la totalidad del acervo teórico disponible, permite deducir resultados nuevos, supuesto que la combinación y el manejo matemáticos de los resultados previamente aceptados de diferentes disciplinas de la física teórica, tomados como premisas del razonamiento, sean compatibles con el análisis dimensional. Como la preducción es una implementación del razonamiento deductivo en el contexto de descubrimiento, doy satisfacción a la queja de Medawar sobre la 'debilidad' del deductivismo hipotético.

Por otra parte, contrariamente a inducción y abducción, que son inferencias ampliativas, la preducción constituye una inferencia anticipativa que adelanta o avanza ideas aún no disponibles, que habrán de contrastarse empíricamente. La diferencia fundamental entre ambas formas de inferencia reside en que los resultados preducidos no proceden, ni vienen sugeridos, por los datos empíricos. Antes al contrario provienen deductivamente del acervo teórico disponible. La preducción procede pues a partir de resultados previamente aceptados, eso sí, no necesariamente como verdaderos, del citado acervo teórico, postulados metodológicamente como premisas del procedimiento inferencial.

En definitiva, sostengo que la preducción es la forma como muchas hipótesis fácticas, leyes y modelos teóricos, son anticipados en física. *Preducir* un modelo teórico, una hipótesis fáctica, o incluso una ley teórica, supone generar matemáticamente una ecuación o una serie de ecuaciones relacionadas entre sí, cuyas consecuencias deben poder ajustarse bien con las observaciones.

Abducción y preducción se complementan mutuamente, pues mientras la primera es la práctica de descubrimiento propia de las ciencias observacionales de la naturaleza, la preducción constituye una estrategia de descubrimiento ampliamente usada en las ciencias teóricas de la naturaleza. De forma que es perfectamente legítimo postular una *tesis de complementariedad* entre abducción y preducción, entre inferencias ampliativas y anticipativas.

Tesis concretas acerca del método científico

Afirmo en primer lugar la debilidad de la idea de la existencia de *un* método científico, sobre todo si esta idea va asociada a un *fetichismo del método.* Pues éste

- condena a las ciencias humanas y sociales a asimilar los métodos de la física,
- bien declara como no científicas determinadas prácticas como la biología evolucionista darwiniana, y por extensión a las ciencias observacionales de la Naturaleza.

Además, el fetichismo del método se compromete con un descuido, incluso una condena, del contexto de descubrimiento científico, lo que constituye uno de los mayores errores de los filósofos de la ciencia de la segunda mitad del siglo XX.

Mi segunda tesis consiste en que el contexto de descubrimiento no sólo incluye a la inducción, entendida como inferencia conservadora de la verdad y ampliadora del contenido. Incluye también a la abducción. Sin olvidar otras formas de creatividad como la serendipia o la analogía, que son prácticas heurísticas fértiles que encauzan la creatividad en ciencia.

Mi tercera tesis es que hay además otra forma del *ars inveniendi* en ciencia, que denomino *preducción*, ampliamente extendida en la metodología de las ciencias físicas, aunque los filósofos de la ciencia no se hayan percatado de ello:

- La *preducción* no es sino la implementación del razonamiento deductivo en el contexto de descubrimiento científico, más allá de su uso en el contexto de justificación o en el de explicación teórica.

Mi cuarta tesis es la de la *complementariedad entre abducción y preducción*. En efecto,

- como la abducción se aplica ampliamente en las ciencias observaciones de la naturaleza: geofísica, paleontología, etc,
- mientras que la preducción se aplica en las ciencias teóricas de la naturaleza: física matemática,
- resulta que ambas se complementan mutuamente, cubriendo ampliamente el espectro de prácticas de descubrimiento en las ciencias de la naturaleza.
- Finalmente, y en particular en ciencias físicas, abducción y preducción son ambas prácticas comunes de descubrimiento.

La *preducción* en el contexto de descubrimiento de ciencias naturales teóricas

A fin de mostrar la presencia de la preducción en los procesos de creatividad en ciencias teóricas de la naturaleza, voy a proceder a su presentación en los pasos siguientes:

1. En aras de la argumentación parto de la idea de Peirce (CP, 5.145) de que la inducción comparte con la deducción el que "nunca puede dar origen a una idea. Tampoco lo hace la deducción".
2. Discrepo de la idea Peirceana de que "Todas las ideas de la ciencia llegan a ella por Abducción." (CP, 5.145)
3. Mi objetivo es responder a la pregunta: ¿*Puede ser usado el razonamiento deductivo en el contexto de descubrimiento científico*? Mi respuesta es simplemente: Sí.
4. Esta respuesta supera la presunta 'debilidad' del método hipotético-deductivo a la que apunta Peter Medawar (1974, p. 289): "The weakness

of the hypothetic-deductive system…lies in its disclaiming any power to explain how hypotheses come into being."

5. Sostengo que en metodología de la física teórica podemos extender el razonamiento deductivo al contexto de descubrimiento, más allá de su uso ordinario como deductivismo axiomático, o como herramienta de comprobación de hipótesis en el contexto de justificación, o en el contexto de explicación.

6. La física teórica se sirve de las matemáticas como herramienta fundamental. El físico teórico puede aplicar a placer el principio leibniziano de *sustitución salva veritate*, donde la salvaguarda de la verdad, o mejor dicho, la legitimidad de las sustituciones realizadas, la garantiza el análisis dimensional. Ello le permite al físico manejar las fórmulas y símbolos de magnitudes a su antojo, y con una única condición: allí donde los resultados de su juego matemático son susceptibles de contrastación empírica, la Naturaleza tiene el veredicto final sobre la *utilidad* o el *interés* del juego.

7. Mi tesis es que entonces resulta reconocible una nueva forma de razonamiento que denomino *preducción teórica* o simplemente *preducción*.

Un ejemplo sencillo de preducción: la de la naturaleza ondulatorio-corpuscular de los fotones. La obtuvo Einstein *combinando dos resultados de dos teorías distintas*, la relatividad especial y la física cuántica de Planck,

a saber: $E=cp$ con $E = h\nu$, de donde resulta $\lambda = h/p$, ó $p = h/\lambda$, que son las fórmulas que manifiestan el carácter dual de la radiación.

Otros ejemplos ilustrativos de preducción teórica los muestro en Rivadulla (2008, 2009*a* y 2009*b*).

8. La *preducción* es pues una forma de razonamiento que parte de resultados previamente aceptados del acervo teórico, postulados *metodológicamente* como premisas del procedimiento inferencial. Estas premisas pueden proceder de diferentes teorías. Y cualquier producto aceptado puede servir de premisa. En el bien entendido que *aceptado* no implica *aceptado en cuanto verdadero.*

9. Esto evoca la noción de método hipotético-deductivo. La *preducción* es efectivamente una implementación del razonamiento deductivo. Su especificidad reside en que extiende el método deductivo al contexto de producción de hipótesis (contexto de descubrimiento).

10. Reservo pues el término *preducción* para la forma de razonamiento deductivo en el contexto de descubrimiento, que consiste en recurrir a resultados aceptados, pertinentes al caso, del conjunto de la física, a fin de anticipar resultados nuevos por combinación y manipulación matemática de aquéllos de forma compatible con el análisis dimensional.

Tesis generales acerca de las relaciones entre abducción y preducción

1. La *preducción* es una forma de razonamiento que parte de resultados previamente aceptados, si bien no asumidos necesariamente como verdaderos, postulados *metodológicamente* como premisas de posteriores inferencias científicas. Estas premisas pueden proceder de diferentes teorías.
2. El razonamiento preductivo difiere del razonamiento *ampliativo*, inductivo o abductivo, en que las hipótesis preducidas no son sugeridas por los datos, sino que su postulación procede deductivamente a partir del acervo teórico disponible.
3. La preducción es entonces una implementación del razonamiento deductivo en el contexto de descubrimiento científico.
4. La preducción proporciona el procedimiento por el que hipótesis fácticas, leyes teóricas y modelos teóricos pueden ser propuestos en física, combinando, de forma compatible con el análisis dimensional, resultados aceptados de teorías disponibles. La preducción es pues una forma de razonamiento *anticipativo.*
5. *Tesis de complementariedad* entre abducción y preducción: Mientras la abducción es la forma preferida de razonamiento en las ciencias naturales observacionales, la preducción lo es, si bien no es la única, en el contexto de descubrimiento de las ciencias teóricas. De manera que ambas cubren ampliamente el espectro creativo en las ciencias de la Naturaleza.
6. *Falibilidad intrínseca de las inferencias abductivas y preductivas*: En el caso de la abducción nuevos datos pueden aparecer en detrimento de las hipótesis propuestas, lo que invita a revisarlas o sustituirlas por otras nuevas, las cuales deben ser compatibles tanto con los datos antiguos como con los nuevos.

   Como reconoce el propio Peirce (CP, 2.777): "La hipótesis que [la abducción, A.R.] concluye problemáticamente, con frecuencia es totalmente falsa, y el procedimiento no necesita llevarnos siempre a la verdad."

   En el caso de la preducción, las hipótesis, modelos y demás resultados preducidos, dependen de la totalidad del acervo teórico disponible, aunque no aceptado en cuanto verdadero. La preducción hace posible la obtención de hipótesis que nos permiten manejarnos predictivamente de modo falible con la Naturaleza.
7. (Inducción), abducción y preducción son estrategias de razonamiento falibles para nuestro manejo científico con la Naturaleza.

El método científico en el contexto de descubrimiento

De lo dicho anteriormente se desprende que no hay un único método científico seguro, sino muchos métodos falibles, adaptados a disciplinas concretas. O, por

mejor decir, *lo que hay en el contexto de descubrimiento son diferentes prácticas o estrategias que sirven de vehículo a la creatividad científica.* Inducción, abducción y preducción son una muestra. Uno pues mi voz a la de aquellos filósofos de la ciencia, aludidos por Howard Sankey (2008, p. 249), dispuestos a reservar un papel al método en el contexto de descubrimiento.

Referencias bibliográficas

Medawar, P. (1974), 'Hypotheses and Imagination', en Schilpp, P. A. (ed.), *The Philosophy of Karl Popper*, La Salle, Ill., Open Court.

Peirce, C. S., (1965), *Collected Papers*, Cambridge, MA, Harvard University Press.

Rivadulla, A. (2008), 'Discovery Practices in Natural Sciences: From Analogy to Preduction', *Revista de Filosofía* 33 (1), pp. 117-137.

— (2009a), 'Anticipative Preduction, Sophisticated Abduction and Theoretical Explanations in the Methodology of Physics', en González Recio, J. L. (ed.), *Philosophical Essays on Physics and Biology*, Hildesheim, Olms.

— (2009b): 'Ampliative and Anticipative Inferences in Scientific Discovery: Induction, Abduction and Preduction', en Fernández Moreno, L. (ed.), *Language, Nature and Science: New Perspectives*, Madrid, México D. F., Plaza y Valdés.

Sankey, H. (2008), 'Scientific Method', en Psillos, S. y Curd, M. (eds.), *The Routledge Companion to Philosophy of Science*, London, Routledge.

El concepto de función biológica desde un enfoque organizacional

Cristian Saborido, Matteo Mossio y Alvaro Moreno
Universidad del País Vasco / Euskal Herria Unibertsitatea
cristian.saborido@ehu.es

En las últimas décadas el concepto de función ha sido objeto de un intenso debate en filosofía de la ciencia y, particularmente, dentro de la filosofía de la biología (cfr. Ariew *et al.* 2002; Buller 1999; Allen *et al.* 1997). La razón fundamental de este interés reside en el hecho de que el concepto de función parece conllevar unas dimensiones *teleológica* y *normativa* que no parecen acomodarse fácilmente dentro de los esquemas clásicos de explicación científica. Por un lado, la atribución de una función a un rasgo introduce aparentemente una "teleología", pues aparentemente esta función contribuye a explicar la misma existencia del rasgo, subvirtiendo de este modo el orden entre causa y efecto (Buller 1999:1-7). Por otro lado, el concepto de función posee una dimensión normativa, ya que hace referencia a (al menos) un efecto que el rasgo analizado *debe* llevar a cabo (Hardcastle 2002: 144).

En la discusión filosófica sobre la función biológica se han presentado una gran variedad de propuestas para dar cuenta de este concepto de un modo científicamente aceptable. Generalmente, estas distintas teorías suelen agruparse en dos perspectivas principales: la etiológica y la disposicional.

El enfoque predominante es el llamado "etiológico", cuya primera formulación fue dada por Wright (1973). Este enfoque define la función de un rasgo apelando a su etiología, esto es, a su historia causal. Las funciones de un rasgo se corresponden con sus efectos pasados que explican causalmente la presencia actual del rasgo. En su versión más extendida, el enfoque etiológico se sustenta en un proceso causal histórico-selectivo, y define la función de un rasgo como aquellos efectos gracias a los cuales los cuales los predecesores de este rasgo fueron mantenidos por la selección natural (Godfrey-Smith 1994; Millikan 1989; Neander 1991). Este enfoque etiológico ofrece así una interpretación aceptable y elegante de la dimensión teleológica, pues explica la existencia de un rasgo individual como la consecuencia de los efectos de los predecesores de este rasgo. Al mismo tiempo, nos da una justificación de la normatividad: los rasgos funcionales han de producir por los cuales han sido seleccionados. La elegancia de la interpretación etiológica es indisociable de su carácter histórico, en la medida en que las atribuciones funcionales no conciernen a la actividad *actual* de un rasgo de un organismo, sino al hecho de que éste posea una determinada historia selectiva. Esto tiene implicaciones problemáticas para las teorías que han sido señaladas por diversos autores (Boorse 1976; Cummins 2002; Davies 2000).

La otra tradición importante, que aquí denominaremos "disposicional", reagrupa a un conjunto muy variado de teorías inspiradas en el análisis desarrollado por Nagel (1977), como la «*Causal Role Theory*» (Cummins 1975; Craver 2001; Davies 2001), la «*Goal Contribution Approach*» (Adams 1979; Boorse 2002, 1976) o la «*Propensity View*» (Bigelow and Pargetter 1987; Canfield 1964; Ruse 1971). Todas estas teorías, a pesar de tener diferencias teóricas más que considerables, comparten un mismo fondo teórico común consistente en rechazar la dimensión teleológica como un elemento constitutivo del concepto de función. Según estas teorías, las funciones de un rasgo no explican su existencia. Las funciones simplemente constituyen una clase particular de efectos o "disposiciones" producidos por un rasgo y que contribuyen a alguna capacidad distintiva del organismo. La discusión en el seno de la tradición disposicional se focaliza en la forma en que se debe definir y restringir las contribuciones y las capacidades por las que las atribuciones funcionales parecen adecuadas, y si se ha de buscar una justificación de lo que un rasgo funcional "debe hacer", es decir, de su normatividad. En contraste con la naturaleza histórica de las teorías etiológicas, la perspectiva disposicional se concentra en los organismos presentes, en un intento de comprender cómo el lenguaje funcional describe una clase específica de relaciones causales de la actividad del sistema que se analiza. De cualquier forma, las teorías disposicionales han sido duramente criticadas por no haber sido capaces de construir una definición suficientemente restringida de función que sirva para determinar los casos en los que las atribuciones funcionales son informativas y pertinentes (Millikan 1989 ; Bedau 1992 ; McLaughlin 2001).

Así pues, el estado actual del debate filosófico nos sitúa ante un dilema con dos soluciones, ninguna de las cuales parece ser muy satisfactoria. Por un lado tenemos las teorías etiológicas, irremediablemente históricas e incapaces de justificar cómo las atribuciones funcionales pueden referirse a propiedades presentes en sistemas biológicos actuales. Por otro lado, las teorías disposicionales son demasiado abarcadoras como para dar una definición útil y correcta de función biológica, no tienen en cuenta la dimensión teleológica y, además, parecen incapaces de fundamentar adecuadamente la dimensión normativa.

Recientemente se han propuesto distintas teorías que pretenden ofrecer una naturalización de los aspectos teleológicos y normativos de las explicaciones funcionales desde una perspectiva que considera los aspectos organizativos de los sistemas vivientes. A través del análisis de las propiedades de la organización actual de los sistemas biológicos, estas teorías organizacionales intentan explicar la dimensión normativa y la dimensión teleológica del concepto de función biológica (McLaughlin 2001, Schlosser 1998, Delancey 2006, Bickhard 2000, Christensen & Bickhard 2002, Collier 2000). Nuestra propuesta pretende ser una nueva contribución dentro de esta línea de trabajo.

En esta comunicación, presentaremos un enfoque organizacional de las funciones, que busca fundamentar sus dimensiones teleológicas y normativas interpretando las funciones como una clase particular de relaciones causales que forman parte de la organización propia de los sistemas biológicos.

Particularmente, este enfoque organizacional parte de la idea de que las funciones se encuentran intrínsecamente ligadas a dos propiedades constitutivas de los sistemas biológicos: el *cierre* y la *diferenciación* organizacional.

Dentro de esta perspectiva el fundamento epistemológico del concepto de función reside en la existencia de una clase de sistemas en los cuales la actividad es orientada, al menos parcialmente, a la preservación de su propia identidad y coherencia (Rosen 1991, Sommerhoff 1959). En particular, la función (o funciones) de un proceso (o, más generalmente, de una relación causal) hace referencia a la producción de un efecto del cual depende, de manera más o menos directa, la existencia del sistema que produce ese mismo proceso. Así, la organización de los sistemas biológicos posee en sí misma las propiedades necesarias que permiten la generación de la teleología y la normatividad constitutivas de las funciones biológicas. Esta perspectiva se sustenta en un marco teórico –desarrollado en las últimas décadas en los dominios de la biología teórica y la teoría de sistemas complejos- que comprende la organización de los sistemas biológicos en término de su *auto-mantenimiento*. De esta forma, interpretamos los sistemas biológicos como una clase específica de sistemas auto-mantenidos, caracterizados por dos propiedades constitutivas: el *cierre* y la *diferenciación* organizacionales. Los sistemas biológicos, en tanto que sistemas cerrados y diferenciados, poseen las propiedades necesarias para fundamentar las atribuciones funcionales a los componentes de su organización constitutiva.

En nuestra teoría (Mossio, Saborido & Moreno, *en prensa*), un rasgo tiene una función si y sólo si está sometido a cierre organizacional dentro de un sistema auto-mantenido diferenciado.

Esta definición implica satisfacer tres condiciones diferentes. Así, un rasgo T tiene una función si y sólo si:

C1: T contribuye al mantenimiento de la organización O de S;

C2: T es producido y mantenido bajo constricciones ejercidas por O;

C3: S es organizacionalmente diferenciado.

Por ejemplo, el corazón tiene la función de bombear la sangre porque (C1) bombear sangre contribuye al mantenimiento del organismo permitiendo la circulación de la sangre, lo que a su vez permite llevar los nutrientes a las celular y evacuar los desechos, estabilizar la temperatura y el Ph, etc. Al mismo tiempo (C2), el corazón está producido y mantenido por el organismo, pues la integridad global es una condición para la existencia del corazón. Por último (C3), el organismo es diferenciado, ya que produce un gran número de estructuras que contribuyen de manera diferente al mantenimiento del sistema.

En nuestra propuesta, la definición organizacional *reemplaza* a las definiciones disposicionales y etiológicas, pues da cuenta al mismo tiempo de la teleología y la normatividad de las funciones desde una perspectiva no-histórica. En efecto, C1 requiere que el rasgo funcional contribuya al mantenimiento de la organización actual, y C2 impone que el rasgo sea generado por la misma organización a la que

contribuye. Consecuentemente, esta definición da cuenta tanto de lo que un rasgo debe hacer como de su misma existencia en el seno del sistema.

Sostenemos que este enfoque fundamenta las dimensiones teleológica y normativa de las funciones en la organización actual de los organismos, en la medida en que nos ofrece una explicación no arbitraria tanto de la existencia del rasgo funcional como de las normas a las que los rasgos funcionales han de obedecer. Sugerimos que, debido a estas propiedades, el enfoque organizacional puede entenderse como una combinación de las perspectivas etiológica y disposicional dentro de un mismo marco teórico unificado.

Referencias bibliográficas

Adams, F. R. (1979), 'A goal-state theory of function attributions', *Canadian Journal of Philosophy*, 9, pp. 493–518.

Allen, C., Bekoff, M. y Lauder, G. V. (eds.) (1998), *Nature's Purposes*, Cambridge, MA, MIT Press.

Ariew, A. R., Cummins, R. y Perlman, M. (eds.) (2002), *Functions,* Oxford, Oxford University Press.

Bedau, M. A. (1992), 'Goal-Directed Systems and the Good', *The Monist,* 75, pp. 34-49.

Bickhard, M. H. (2000), 'Autonomy, Function, and Representation', *Communication and Cognition - Artificial Intelligence*, 17, pp. 3-4, 111-131.

Bigelow, J. y Pargetter R. (1987), 'Functions', *Journal of Philosophy* 84, pp. 181-196.

Boorse, C. (2002), 'A Rebuttal on Functions', en Ariew, A., R. Cummins y M. Perlman (eds.), *Functions*, Oxford, Oxford University Press, pp. 63–112.

Buller, D. J. (1999), *Function, Selection, and Design*,Albany, New York, SUNY Press.

Canfield, J. (1964), 'Teleological explanation in biology', *British Journal for the Philosophy of Science* 14, pp. 285–295.

Christensen, W. D. y Bickhard, M. H. (2002), 'The Process Dynamics of Normative Function', *The Monist* 85, pp. 1, 3-28.

Cummins, R. (1975), 'Functional analysis', *Journal of Philosophy* 72, pp. 741-65.

— (2002), 'Neo-Teleology', en Ariew, A., Cummins, R. y Perlman, M. (eds), *Functions,* Oxford, Oxford University Press, pp. 157-172.

Collier, J. (2000), 'Autonomy and process closure as the basis for functionality', en G. Chandler y J. L. R van de Vijver (eds.), *Closure: Emergent Organizations and their Dynamics*, volume 901 of the New York Academy of Sciences.

Craver, C. F. (2001), 'Role functions, mechanisms, and hierarchy', *Philosophy of Science* 68, pp. 53–74.

Davies, P.S. (2001), *Norms of Nature. Naturalism and the Nature of Functions*, Cambridge, MIT Press.

Delancey, C. (2006), 'Ontology and teleofunctions: A defense and revision of the systematic account of teleological explanation', *Synthese* 150, pp. 69-98.

Edin, B. (2008), 'Assigning biological functions: making sense of causal chains' *Synthese* 161, pp. 203-218.

Godfrey-Smith, P. (1994), 'A modern history theory of functions', *Noûs* 28, pp. 344-362.

Hardcastle, V. G. (2002), 'On the normativity of functions', en Ariew, A., Cummins, R. y Perlman, M. (eds), *Functions,* Oxford, Oxford University Press, pp. 144-156.

McLaughlin, P. (2001), *What Functions Explain. Functional Explanation and Self-Reproducing Systems*, Cambridge, Cambridge University Press.

Nagel, E. (1977), 'Teleology revisited', *Journal of Philosophy* 74, pp. 261-301.

Neander, K. (1991), 'Functions as selected effects: The conceptual analyst's defense', *Philosophy of Science* 58, pp. 168-184.

Millikan, R. G. (1989), 'In defense of proper functions', *Philosophy of Science* 56, pp. 288–302.

Mossio, M., Saborido, C. Moreno, A. (en prensa), 'An Organizational Account for Biological Functions', *British Journal for the Philosophy of Science*, <http://bjps.oxfordjournals.org/cgi/content/abstract/axp036>.

Mossio, M., Saborido, C. y Moreno, A. (en prensa), 'Fonctions: Normativite, Teleologie et Organisation', en Gayon, J. y De Ricqles, A. (eds), *Epistemologie de la categorie de fonction: des sciences de la vie a la technologie*, Paris, P.U.F. <http://www.ehu.es/ias-research/doc/Mossio_Saborido_Moreno_PUF.pdf>.

Ruse, M. (1971), 'Functional statements in biology', *Philosophy of Science* 38, pp. 87-95.

Rosen, R. (1991), *Life itself: A comprehensive inquiry into the nature, origin and fabrication of life*, New York, Columbia University Press.

Schlosser, G. (1998), 'Self-re-production and functionality: A systems-theoretical approach to teleological explanation', *Synthese* 116, pp. 303–354.

Sommerhoff, G. (1959), 'The abstract characteristics of living organisms', en Emery, F. E. (ed.), *Systems Thinking*, London, Harmondsworth.

Wright, L. (1973), 'Functions', *Philosophical Review* 82, pp. 139-168.

Zahar y Feyerabend: dos nociones de "equivalencia observacional"*

Fernanda Samaniego
Universidad Complutense de Madrid
Fernanda.Samaniego@gmail.com

La transición de la teoría de Lorentz a la teoría de la Relatividad fue explicada por Elie Zahar (1973) en términos lakatosianos. Dada la "equivalencia observacional" que existía entre ambos programas de investigación antes de 1915, argumenta Zahar, es la superioridad heurística del programa de Einstein lo que nos permite entender el abandono de la teoría del éter y la aceptación de la Relatividad por parte de algunos científicos entre 1904 y 1914. A pesar de su agudeza y minuciosidad, la propuesta de Zahar ha recibido críticas severas. Aquí nos enfocamos en una de ellas (Feyerabend 1974), según la cual, la tesis de la equivalencia observacional entre las dos teorías es falsa. El objetivo del presente trabajo es mostrar que los autores manejan dos nociones distintas de equivalencia observacional. A la luz de este resultado se argumenta que sólo una versión matizada y debilitada de la crítica de Feyerabend puede defenderse.

Más allá de los datos empíricos

Cuando los resultados empíricos no son suficientes para elegir entre dos programas de investigación en competencia, necesitamos criterios extra empíricos para explicar que la comunidad científica favorezca a uno de los dos programas y abandone el otro. Una situación de este tipo se dio entre 1904 y 1914 respecto al programa del éter de Lorentz (L) y el programa de la Relatividad de Einstein (R). En 1904 Lorentz ya había presentado la versión definitiva de su teoría, conocida como Teoría de los Estados Correspondientes. En 1915, según Zahar, se da la sustitución empírica de L por R con la explicación del comportamiento anómalo de Mercurio. Pero durante los años comprendidos entre estos dos sucesos ya había muchos adeptos al programa de Einstein; cabe entonces cuestionarnos: ¿Qué tenía la propuesta de Einstein para ser capaz de convencer a un científico a abandonar el programa de Lorentz? Si ambas teorías eran equivalentes observacionalmente durante ese periodo ¿por qué algunos científicos prefirieron la contra-intuitiva Relatividad frente a la teoría del éter?

En su análisis lakatosiano Zahar argumenta que la superioridad del poder heurístico del programa de Einstein fue lo que le permitió imponerse sobre el programa de Lorentz. La heurística de L se basaba en el principio de que "*todos*

* Este trabajo se realizó con financiamiento del Consejo Nacional de Ciencia y Tecnología y surgió en el marco del curso de doctorado "Filosofía de la Física" impartido por el Dr. Mauricio Suárez Aller en la Universidad Complutense de Madrid.

los fenómenos físicos están gobernados por acciones transmitidas por el éter" [Zahar (1973a), p.100], mientras que la heurística de R consistía en la búsqueda de simetrías tanto a nivel observacional como a nivel ontológico.

Es esencial para el problema que nos ocupa notar que la "equivalencia observacional" entre el programa de Lorentz y el de Einstein no es una tesis "implícita" o simplemente "presupuesta" en el análisis de Zahar. Por el contrario, es una tesis que Zahar defiende explícita y cautelosamente, cuando afirma que:

> "El significado filosófico de la Teoría de los Estados Correspondientes es que podía ser, como Poincaré mostró, fácilmente transformada en una teoría equivalente observacionalmente a la Relatividad Especial. Entonces, para comparar los méritos de las dos teorías rivales en el año 1905, debían invocarse criterios no-empíricos." [Zahar(1973), p.116]

También es importante resaltar el estatus peculiar que Zahar asigna al éter, al tomarlo como un elemento de la heurística positiva y no incluirlo en el núcleo del programa de Lorentz. Este es sin duda uno de los puntos más controversiales de la propuesta de Zahar, a quien se le ha sugerido corregir su caracterización del núcleo lorentziano.

Críticas y evolución de la propuesta de Zahar

Las críticas a la reconstrucción del caso Lorentz-Einstein de Zahar pueden dividirse en tres grupos:

i) Objeciones a la caracterización de los programas de investigación R y L. Se sugiere sustituir en el núcleo de L las ecuaciones de Maxwell por las ecuaciones microscópicas del propio Lorentz [Schaffner 1974]; Incluir al éter en el núcleo de L y excluir la 3ª Ley de Newton por ser incompatible con las ecuaciones de la teoría electromagnética [Miller, 1974]. Por otro lado, se considera que el núcleo del programa de Einstein debe incluir la teoría del átomo einsteniana, con lo cual ya no podemos decir que L fue sustituida únicamente por R, sino que derivó en dos programas distintos e incompatibles: la Teoría de la Relatividad y la Mecánica Cuántica [Feyerabend, 1974]. A este primer tipo de críticas Zahar respondió con mucha flexibilidad, aceptando los cambios y haciendo hincapié en que esto no afecta a su tesis de la superioridad heurística del programa de Einstein.

ii) Factores históricos que fueron omitidos o que merecen un papel más importante. La influencia de Hume o de Mach, la crisis de la Mecánica Cuántica, el grado de simplicidad de R comparado con el de L, el concepto espacio-tiempo y los resultados negativos del experimento Michelson-Morley, son algunos de los factores a los que Zahar no asigna suficiente importancia según sus críticos [Shafner (1974) y Stachel (1989)].

iii) Cuestionamiento de los presupuestos o definiciones filosóficas de Zahar. Se argumenta que Zahar utiliza un criterio presentista de "objetividad" aplicado a las decisiones "racionales" de los científicos. El hablar de decisiones "objetivas" en un momento dado va en contra de la metodología misma de Lakatos, cuyo rango de aplicación está conformado por programas de investigación que ya han evolucionado durante largos periodos de tiempo y han sufrido por tanto "cambios

progresivos" o "cambios degenerativos" [Feyerabend, 1974]. Dentro de este tercer grupo de críticas, se encuentra la crítica al presupuesto de la equivalencia observacional entre R y L. Esta es la crítica que veremos aquí con más detalle y la considero de especial relevancia ya que, de ser correcta, todo análisis como el de Zahar carecería de sentido.

La crítica a la tesis de la equivalencia observacional

Cuando dos programas en competencia son capaces de predecir lo mismo una manera de decidirse por uno de ellos es elegir el que menos elementos utilice para dar cuenta de lo mismo–navaja de Ockham--. Feyerabend formula este criterio de la siguiente manera: "…*un programa de investigación que contiene un elemento sin función alguna en un momento dado t, es en ese momento inferior a un programa de investigación que carece del elemento pero genera exactamente las mismas predicciones*" [Feyerabend (1874), p.28]. Según Feyerabend, éste es el criterio que utiliza Zahar para argumentar que el éter era un elemento inútil en L y que, por ende, L era inferior a R. Pero para Feyerabend, al utilizar dicho criterio, Zahar asume falsamente que L y R eran equivalentes observacionalmente.

El problema que quiero señalar aquí consiste en que la noción de "equivalencia observacional" que Feyerabend está utilizando es distinta a la de Zahar. Cuando Zahar habla de "equivalencia observacional" se refiere únicamente a las observaciones que ya se habían hecho *de facto* en un momento dado. En la siguiente cita podemos ver que Feyerabend claramente maneja una noción diferente, según la cual, incluso los hechos que no han sido observados son relevantes para decidir la equivalencia observacional:

> "…él [Zahar] asume que L y R (1905) eran observacionalmente equivalentes. Esto no es verdad. Incluso en 1905 (o en 1906, 1907,...) la equivalencia sólo se había establecido para estados de equilibrio: si las contracciones y las dilataciones de longitud, de tiempo, los cambios de masa, son resultados de la interacción entre el éter y la materia circundada por él, entonces cualquier cambio de velocidad relativa al éter llevará a oscilaciones. La Relatividad, por su parte, no da cuenta dichas oscilaciones" [Feyerabend (1974), p.28]

Nociones de "equivalencia observacional"

Un primer punto a señalar es que R no excluía al éter. El programa einsteniano R podría mantenerse intacto tomando uno de los marcos inerciales y llamándolo "marco del éter". Pero concedamos por un momento que R excluía al éter y concentrémonos en las nociones de equivalencia observacional.

Ciertamente, podemos decir que L predice un hecho que no se había detectado --a saber, las oscilaciones producidas por un objeto acelerándose respecto al éter-- acerca del cual R no dice nada, porque su campo de dominio son los marcos de referencia inerciales. Si queremos argumentar, como hace Feyerabend, que de esto se sigue que R y L no son equivalentes observacionalmente, entonces estaríamos sosteniendo que:

> T y T' son equivalentes observacionalmente sii sus predicciones observacionales son iguales.

Reichenbach, Putnam, Salmon y Sklar han manejado nociones similares a ésta[1]. Lo que resulta curioso es que Feyerabend recurra a ondulaciones del éter, cuando tenemos un ejemplo mucho más obvio de que L y R tienen consecuencias observacionales distintas: la predicción que R hace de la influencia de la masa solar sobre la trayectoria de Mercurio. Queda claro entonces que L y R tienen predicciones observacionales distintas, ¿de esto sigue acaso que la equivalencia observacional de L y R es falsa? Para Feyerabend sí y para Zahar no, porque la respuesta depende de la noción que se defienda de equivalencia observacional.

Lo que Zahar tiene en mente es que L y R eran equivalentes dadas las observaciones que *de facto* se habían realizado hasta 1905, 1906... 1914. Quizás habría sido mucho más afortunado sostener que eran "observacionalmente indiscernibles". *De facto* lo eran, pero potencialmente no ya que existían algunas consecuencias de L y de R que, de ser observadas, favorecerían a uno de los dos programas. De hecho, como Feyerabend mismo menciona [ver Feyerabend (1974), p.28], fue hasta 1936 que los experimentos de Wood, Tomilson y Essen sometieron a prueba las consecuencias observacionales de L de manera definitiva.

Reformulemos la crítica de Feyerabend utilizando la noción de equivalencia observacional como "equivalencia en la capacidad de dar cuenta de un dominio dado"[2]. Esta noción es neutral respecto al debate en tanto no menciona si los hechos han sido observados o no. Las ondulaciones en el éter no forman parte del dominio de R. Por tanto, L y R no son equivalentes observacionalmente bajo esta nueva noción y Feyerabend estaría en lo correcto. Su crítica puede formularse como la tesis matizada:

> L y R *en principio* podían diferenciarse observacionalmente.

Conclusiones

La tesis de la equivalencia entre R y L que defiende Zahar, es legítima bajo su propia noción de "equivalencia observacional".

No obstante, L y R no eran equivalentes en tanto había algunas consecuencias de L y de R que, de ser observadas, favorecerían a uno de los dos programas. Para defender esta tesis –"tesis matizada"- basta referirse a las predicciones sobre el perihelio de Mercurio. No es necesario recurrir a las oscilaciones provocadas por aceleraciones respecto al éter, como hace Feyerabend en su crítica, predicción que no forma parte del dominio de la Relatividad. Zahar sin duda apoyaría también esta tesis matizada.

---

[1] T y T' son e.o. si Γ y Γ' son idénticas, donde Γ contiene sólo los teoremas de T que son contrastables mediante la observación directa; T y T' son e.o. si la evidencia para una es evidencia para la otra. [Ver Glymour (1970), p.276]

[2] Michael Gardner define de esta manera la "equivalencia empírica" para evitar el término "observable", que resulta sumamente problemático en el caso de la mecánica cuántica. Aquí haremos uso de su definición con un objetivo distinto.

Referencias bibliográficas

Feyerabend, P. K. (1974), 'Zahar on Einstein', *The British Journal for the Philosophy of Science* 25, pp. 25-28.

Gardner, M. (1976), 'The Unintelligibility of 'Observational Equivalence'', *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1976, pp. 104-16.

Glymour, C. (1970), 'Theoretical Realism and Theoretical Equivalence', *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1970, pp. 275-88.

Miller, A. I. (1974), 'On Lorent'z Methodology', *The British Journal for the Philosophy of Science* 25, pp. 29-45.

Schaffner, K. F. (1974), 'Einstein versus Lorentz: Research Programes and the Logic of Comparative Evaluation', *The British Journal for the Philosophy of Science* 25, pp. 45-78.

Zahar, E. (1973a), 'Why Did Einstein's Programme Supersede Lorentz's? (I)', *The British Journal for the Philosophy of Science* 24, pp. 95-123.

— (1973b), 'Why Did Einstein's Programme Supersede Lorentz's? (II)', *The British Journal for the Philosophy of Science* 24, pp. 223-62.

— (1977), 'Mach, Einstein, and the Rise of Modern Science', *The British Journal for the Philosophy of Science* 38, pp. 195-213.

— (1978), 'Einstein's Debt to Lorentz: A Reply to Feyerabend and Miller', *The British Journal for the Philosophy of Science* 29, pp. 49-60.

— (1983), *Einstein's Revolution; A study in heuristics*, La Salle, Illinois, Open Court.

Common causes, measurement dependence and no-conspiracy: ontological implications*

Iñaki San Pedro
CPNSS, London School of Economics
i.san-pedro@lse.ac.uk

Introduction

It is still an open question whether common cause explanations are appropriate to account for EPR correlations. The received view takes it that this kind of explanation is to be ruled out. Typically, such arguments take common causes as hidden variables onto which several constraints and restrictions are set. Constraints on the common causes are intended to reflect standard requirements typical of any physical system, including temporal order of causal relations or locality considerations. As a result, some version of the Bell inequalities is derived. The strength of the argument relies on the plausibility of the conditions imposed on the common causes.

One such condition is the so-called "no-conspiracy" condition. It reflects the requirement that the postulated common causes be independent of the measurement settings. A violation of such independence is usually taken to entail certain strange conspiratorial behaviour, unless backwards in time causation is brought into the picture. This interpretation presupposes that the events corresponding to the common causes take place prior to measurement.

This standard reading of violations of "no-conspiracy"-like conditions may be challenged if common causes are postulated to take place upon, or right after, measurement operations (San Pedro, forthcoming). Violation of "no-conspiracy" need not be a consequence of a world conspiracy in this case, nor to causes operating backwards in time. To the contrary, it just responds to the fact that the postulated common causes are *measurement dependent.*

The object of this paper is to provide an account of the implications of the violation of "no-conspiracy" conditions and evaluate the possible ontological consequences.

* Research supported by the Basque Regional Government Postdoctoral Fellowship Programme (Programa de Perfeccionamiento de Doctores en el Extranjero del Departamento de Educación, Universidades e Investigación, Gobierno Vasco) and the Spanish Ministry of Science and Innovation (research project FFI2008-06418-C01-03). I am grateful to Mauricio Suárez for useful comments.

A non-Factorizable Common Cause Model for EPR Correlations

In a previous paper (San Pedro, forthcoming) I suggested a possible common cause model for EPR correlations. The model builds on the intuition that Reichenbachian common causes play fundamentally an explanatory role and include explicit information about the measurements performed. In particular, measurement operations are causally relevant to the postulated common causes. Furthermore, common causes in the model are not viewed as *hidden variables* as such since they are not aimed at completing the formalism of quantum mechanics in any way. Again, their role is fundamentally an explanatory one.

The explicit *measurement dependence* of the common causes is the key feature of the model and is justified by the fact that the postulated common causes take place right after measurement operations are performed. This gives rise to a violation of the "no-conspiracy" condition, i.e.

$$p\left(L_i \wedge C_{ij}^{ab}\right) \neq p\left(L_i\right)p\left(C_{ij}^{ab}\right)$$
$$p\left(R_j \wedge C_{ij}^{ab}\right) \neq p\left(R_j\right)p\left(C_{ij}^{ab}\right)$$

which in turn is responsible for the violation of "factorizability". That is

$$p\left(L_i^a \wedge R_j^b \mid L_i \wedge R_j \wedge C_{ij}^{ab}\right) \neq p\left(L_i^a \mid L_i \wedge C_{ij}^{ab}\right)p\left(R_j^b \mid R_j \wedge C_{ij}^{ab}\right).$$

In the above expressions sub-indices *i* and *j* indicate the directions along which measurement devices can be set. Super-indices *a* and *b* indicate the possible outcomes of an EPR experiment in each direction of measurement. Thus L_i^a represents the event that the outcome of a spin measurement in the left wing of the experiment takes the value *a* (where $a = +$ for spin-up or $a = -$ for spin-down). Similarly for R_j^b. In a similar manner L_i represents the event that measurement in the left wing has been performed along direction *i*. Finally C_{ij}^{ab}, represents the postulated common cause, which includes the labels of the corresponding measurement operations and resulting outcomes.

Violation of "factorizability" in the model is therefore due specifically to the violation of the implicit assumption that common causes need to take place prior to measurement operations. This is indeed a crucial assumption, which is also behind the usual claims regarding free will. In particular, the claim that "no-conspiracy"-type conditions guarantee free will (of the experimenter) makes sense only if the this assumption is in place. However, this assumption seems ultimately unwarranted —it might be challenged, for instance, by looking at EPR correlations from a purely phenomenological perspective (cf. San Pedro, forthcoming). Rejecting it allows us to make sense of *measurement dependence* (of common causes) without appealing to backwards in time causation.

The violation of "no-conspiracy" in the model has diverse consequences, especially as regards the ontology of the event structure.

Ontological Implications I. A non-Local Events Ontology

A first possible interpretation of the model would take the postulated common causes as non-localised events, spreading out in a space and time region. This in turn allows for all causal interactions in the EPR set-up to be local.

The causal process at work may be broken in two steps. In a first stage the common cause is seen as an effect of the measurement operations on both wings, i.e. L_i and R_j and the event Λ associated to the spin singlet state ---which may also include some other causally relevant factors. Each of these operates (causally) *locally* but their conjunction has as a result the non-localised common cause C_{ij}^{ab}, which spreads over space-time. In a second step the non-localised common cause acts again *locally* to produce the corresponding outcome events. The resulting causal structure is represented in the space-time diagram in Figure 1.

Note that it is the conjunction of *both* measurement operations which reveals the non-local character of the common cause event. For it is the conjunction of *both* them that are causally relevant for the common cause. This may also suggest a holistic interpretation of the postulated common cause.

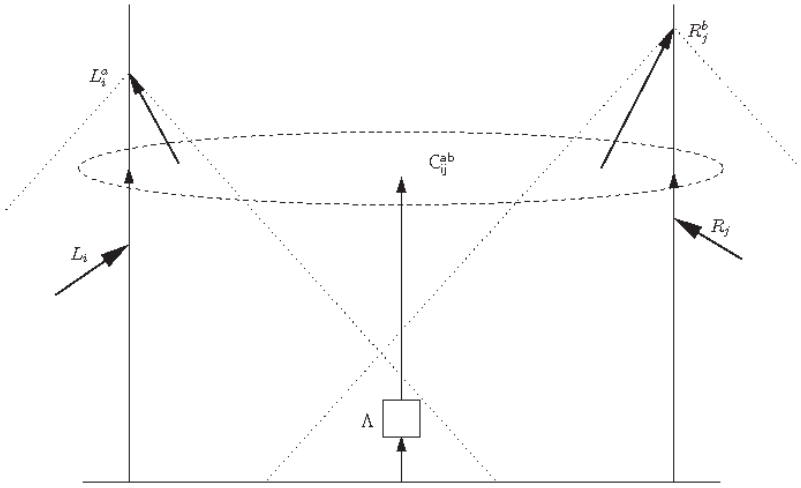


Figure 1: Space-time diagram of the model under the view that common causes are non-localised events acting locally to cause the corresponding EPR outcomes. (Causal influences are represents by solid lines).

The common causes do not necessarily spread out to cover whole slices of the outcome's backward light cones (represented with dotted lines in the diagram). If such were the case, they would *deterministically* cause the corresponding outcomes. But this could result in a conflict with Humean supervenience.[1] For it would entail that the very same causal factors would be responsible for *all*

1 I owe this observation to Mauricio Suárez.

common causes C_{ij}^{ab} $(a,b=+,-;\quad i,j=1,2,3)$ postulated for the diverse correlations in the experiment. But since common causes in the model are assumed to be different to each other, supervenience is violated.[2]

In order to avoid such difficulties, and unless we are prepared to reject Humean supervenience altogether, we need to allow for the common causes to operate in a genuinely indeterministic manner. We may for instance want to require Λ to incorporate different causal factors (besides those directly associated to the spin singlet state) for each of the obtaining common causes.

On the other hand, these non-localised common cause events are fully compatible with relativity. Their structure resembles to that of Teller´s *quanta*, introduced in relation to the notion of fundamental particle in quantum field theory.[3] This is not to say, of course, that our common causes are (or represent) fundamental particles, i.e. excitation states of quantum fields. It seems more appropriate to say that the common cause events "inherit" the typical properties of Teller's *quanta* —electrons, photons and the like— that are involved in EPR experiments.

They will inherit, for instance, the problems related to indistinguishability that *quanta* suffer, precisely due to their space-time non-localizability. We may want to claim, though, that this is not as problematic in our case. For, after all, our common cause events are event types and thus lack already of space-time localizability. Event types seem to avoid therefore the difficulties of *quanta* in this sense. However, it is not less true that the causal relations that we attribute to event types are based on the particular causal relations among the token events that may be taken to constitute them. Thus, the problem of indistinguishability stands.

A possible solution to this is to claim that the identification of *quanta* does not necessarily require that specific space-time locations are provided. *Quanta* may be individually identified for instance by their causal location, i.e. through their causal past and future.[4] This seems certainly even better for our purposes since it gives support to the idea that time priority of common causes with respect to measurement operations is not a fully warranted assumption.

Ontological Implications II. Non-Local Causal Influence

In a second alternative interpretation of the model we may want to retain the more widespread view of events as well localised spatio-temporal entities.[5] By contrast,

2 The assumption that the common causes are different to each other is needed in order to avoid difficulties related to so-called *common*-common causes, whose existence cannot be guaranteed in all cases. See (San Pedro, forthcoming) for details.

3 Cf. (Teller, 1995). Very roughly, *quanta* may be defined as the occurrences of well defined excitation states of the quantum field.

4 This idea of causal identification goes back originally to Davidson's "causal identity" criterion of events. It is (Bartels, 1999) however who applies it to the notion of *quanta*.

5 Again, here I take it that type events are just collections of tokens.

the causal relations between such localised events need to be non-local in order to fit the model's structure. In particular, we seem to be bound to explain our postulated common causes as (space-time localised) events which, on measurement, are capable of exhibiting distant causal powers to produce the corresponding outcomes. Moreover, the common cause events themselves seem to arise as well as effects of other non-local influences. In particular, measurement operations L_i and R_j (one of them at least) need to operate non-locally for them to be causally relevant to the common cause, as the model requires. This is represented in the diagram of Figure 2.

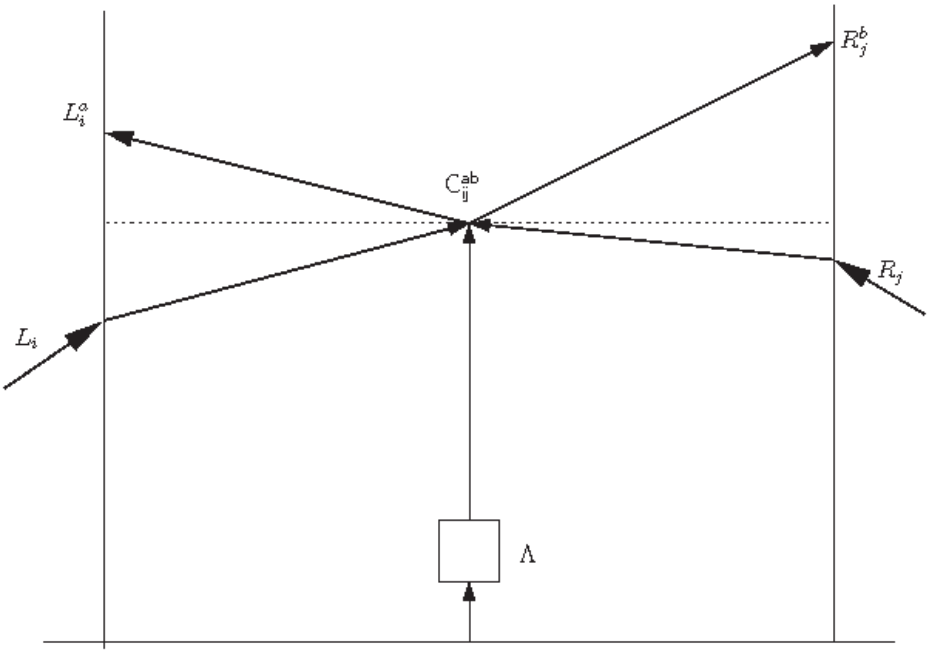


Figure 2: The common causes are space-time localised events which act non-locally to cause the corresponding EPR outcomes. (Causal influences are represented by solid lines).

Unlike in the previous case, the space-time location of the common causes guarantees now their distinguishability. The main problem we are to face with this "non-local causal powers" interpretation is that they seem to operate through superluminal influences, which may be in conflict with relativity theory. The key issue is whether the so-interpreted model's causal structure would allow, not only for superluminal "bare" influences, but also for superluminal signalling. For it is the possibility of superluminal signalling that is really in conflict with relativity theory.

There are at least two possible moves in order to avoid such difficulties. On the one hand we may note that the notion of signalling may be sensitive to the concept of causation we are endorsing. The notion of signalling is usually taken to involve

transfer of some physical quantity (energy, matter, etc.). But if this is so, counterfactual causation, for instance, would not seem to involve signalling even in the case of superluminal causal influences. Alternatively, even if the model allows for superluminal signalling, the existence of entities such as *tachyons* −which propagate superluminally− has not been empirically refuted to date, nor it seems to be in clear conflict with special relativity.[6]

In sum, localised common cause events with non-local causal powers provide as well a conceptually viable interpretation of the causal structure in the model.

Finally I shall point out that it is not my intention to endorse any of the two interpretations of the model at this point. A detailed discussion of the respective merits of both interpretations must await further work.

References

Bartels, A. (1999), 'Objects or Events?: Towards and Ontology for Quantum Field Theory', *Philosophy of Science*, 66, Supplement. *Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association* (Part I: Contributed Papers), S170-84.

Einstein, A., Podolsky, B. and Rosen, N. (1935), 'Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?' *Physical Review* 47, pp. 777-80.

Maudlin, T. (1994), *Quantum Non-Locality and Relativity*, Oxford, Blackwell Publishing, 2nd edition, 2002.

Price, H. (1996), *Time's Arrow and Archimedes' Point,* New York, Oxford University Press.

San Pedro, I. (forthcoming), 'Measurement Dependence is not Conspiracy: A Common Cause Model for EPR Correlations', Submitted to *Studies in History and Philosophy of Modern Physics,* available at <http://arxiv.org/abs/0905.3859>.

Szabó, L. E. (2000), 'On an Attempt to Resolve the EPR-Bell Paradox via Reichenbachian Concept of Common Cause', *International Journal of Theoretical Physics* 39, pp. 911–26.

Teller, P. (1995), *An Interpretative Introduction to Quantum Field Theory*. Princeton, Princeton University Press.

6 (Maudlin, 1994, Ch. 3) provides a nice account of such views.

Datos y Explicación. Dos estrategias complementarias para abordar el problema de Duhem*

Francisco Saurí
Universitat de València
Francisco.Sauri@uv.es

1. Se suele entender por problema de Duhem el que una hipótesis H puede siempre ser defendida ante observaciones que vayan en su contra, en la medida en que reemplacemos los supuestos auxiliares A por una alternativa A' conveniente.

El problema de Duhem conduce a un argumento de infradeterminación. Los argumentos de infradeterminación tienen como conclusión que, dada una teoría sobre inobservables de la que se deducen los hechos observables, siempre hay otras teorías incompatibles de las que se deducen los mismos hechos. La pregunta que plantea el reto de la infradeterminación es: ¿Tenemos conocimiento teórico? Y si es así ¿Cómo lo obtenemos?

Recordemos que la primera condición para confirmar deductivamente una hipótesis H es, en términos lógico - formales, que de la verdad de la hipótesis H y los supuestos auxiliares A (juntos y no por separado) se deducen las observaciones E. En términos probabilísticos p(E/H&A)=1.

Pero lo habitual es que las observaciones también sean deducibles de otras hipótesis lógicamente incompatibles con H. Y entonces, no podemos decir que hemos confirmado una hipótesis H. Por tanto, hay que proponer alguna condición más.

Dos alternativas son:

(1ª) exigir que H explique E

(2ª) exigir que las alternativas a H sean descartables.

Este trabajo pretende comparar ambas alternativas fijándose en cómo aborda cada una el problema de Duhem. En representación de la estrategia de la explicación (la 1ª), se usa un artículo de Weber [Weber (2009)] quien defiende la inferencia de la mejor explicación. El representante elegido para la estrategia (2ª) es el experimentalismo de D. Mayo y su estadística del error [Mayo (1996) y (2006)].

2. Weber expone [Weber (2009), sec. 3] como caso de estudio la investigación sobre la replicación del ADN. Además de la conocida hipótesis de Watson y Crick, denominada en su época hipótesis semiconservadora, se propusieron otras

* La elaboración de este trabajo ha contado con la ayuda del Ministerio de Ciencia e Innovación. Código del proyecto: FFI2008-01169/FISO.

hipótesis. Se consideró que un experimento llevado a cabo por Meselson y Stahl confirmaba la hipótesis semiconservadora. Sin embargo, una de las hipótesis alternativas, la denominada hipótesis conservadora, también era compatible con el experimento de Meselson y Stahl.

Recordemos que actualmente se sabe que la replicación del ADN comienza con su apertura longitudinal y que luego se forman nuevas cadenas de ADN por adición de moléculas en cada una de las mitades longitudinales. La hipótesis conservadora explicaba la formación de las nuevas moléculas por mera copia de la cadena completa sin ninguna división previa de la cadena.

El experimento de Meselson y Stahl partía de una población de *E. Coli* en la que se había sustituido el nitrógeno necesario para la síntesis del DNA por un isótopo más pesado. Luego se trasladaba la población a un medio con el isótopo normal más ligero y se la dejaba reproducirse. Se tomaba una muestra al principio, y luego muestras sucesivas a intervalos regulares de la reproducción de la bacteria. Entonces se sometían las muestras a un proceso de centrifugación en una solución adecuada que situaba el DNA de la *E. Coli* a un nivel de profundidad de la solución según su densidad.

Lo que se observaba es que tras una generación aparecía una banda de densidad intermedia. Tras otra generación, la banda intermedia estaba todavía presente pero había aparecido una nueva banda correspondiente al ADN ligero. Una interpretación obvia de este patrón era que la banda de densidad intermedia era la banda de las moléculas híbridas compuesta de una parte ligera y otra pesada. Las híbridas podían haber sido producidas por el esquema semiconservador, de acuerdo con el cual cada nueva doble hélice preserva un lateral de la molécula de partida. [*ibid.*, p. 27]

Por el contrario, el mecanismo conservador no debía producir una banda de densidad intermedia a menos que se supusiese, como supuesto auxiliar adicional (que podemos llamar A'), que la banda intermedia representaba moléculas-padre de ADN con nitrógeno pesado pegadas a las moléculas-hijo de ADN con nitrógeno ligero.

En otro experimento posterior, Meselson y Rolfe [*ibid.*, sec. 5], demostraron que eso no ocurría. Pese a ello, tal como lo cuenta Weber, los científicos aceptaron la hipótesis semiconservadora *antes* del segundo experimento, es decir, antes del experimento de Meselson y Rolfe

3. El experimentalismo de Mayo exige la siguiente condición añadida para que haya confirmación: la probabilidad de que la hipótesis H supere el procedimiento de contrastación con un resultado tan bueno como E, supuesta la falsedad de H, es muy baja . Es decir: $p(E/\text{no } H) \ll 1$. [Mayo (1996), cap. 6, sec. 2].

Conviene atender el hecho de que Mayo está haciendo hincapié en el proceso de contrastación y, en última instancia, el que éste sea severo quiere decir que sea fiable [Mayo (1996), pp.9, 11-12]. Y esa fiabilidad viene dada porque el procedimiento ha sido utilizado y sabemos qué información nos puede dar. (Mayo

pone esto en términos estadísticos desde una visión frecuencialista de la probabilidad -[Mayo (1996), cap. 11]).

Un procedimiento severo se habría dado en la confirmación de la estructura de doble hélice del ADN mediante los rayos X [Weber (2009) sec. 3], donde el modelo experimental, la conexión entre la evidencia y el mecanismo propuesto, no ofrecía dudas. Qué podíamos o no podíamos saber de una molécula como el ADN mediante los rayos X ya había sido establecido con anterioridad y podía ser utilizado en este caso.

Igualmente, podemos comparar el concepto de procedimiento de contrastación de Mayo con un aparato del que sabemos que es fiable en determinadas situaciones porque ha sido puesto a prueba en esas situaciones. La severidad del proceso de contrastación es la contraparte de la fiabilidad de un aparato.

Para Mayo, el problema de Duhem se bloquea porque las posibles alternativas a la hipótesis son descartadas gracias a la condición de severidad. El procedimiento de contrastación sabemos que dará resultados fiables de la misma manera que sabemos que un aparato funcionará bien si opera en las condiciones de funcionamiento comprobadas.

Y en efecto, Mayo puede señalar (como bien dice el propio Weber [Weber (2009), sec. 5]) que hasta el segundo experimento, el de Meselson y Rolfe, en realidad, no había evidencia confirmadora concluyente a favor de la hipótesis semiconservadora. La severidad del experimento falla en la medida en que la conexión entre la evidencia y el mecanismo propuesto, la conexión entre experimento e hipótesis, no ha sido establecida [Mayo (1996), pp. 147-148]. En concreto, en este caso, el procedimiento de contrastación no es fiable porque el supuesto auxiliar adicional A' no ha sido contrastado. Es como si pusiésemos a funcionar un aparato a 60ºC y nos fiásemos de él cuándo sólo sabemos que funciona bien entre 5ºC y 40ºC.

4. Pero Weber insiste en que ésta no puede ser la respuesta porque entonces "[...] nunca podremos decir que el experimento soportó la hipótesis de Watson - Crick [la hipótesis semiconservadora]; cualesquiera que fuesen las pruebas adicionales que se hiciesen." [Weber (2009), p. 33]. Según Weber, siempre podremos reinterpretar los resultados para plantear el problema de Duhem. Por otra parte, tal como lo cuenta Weber, los científicos no habían realizado una contrastación severa pero, pese a ello, aceptaron los resultados. El porqué, según Weber, es que el mecanismo semiconservador era una buena explicación, y esto es suficiente para la confirmación.

5. Según Weber, la inferencia a la mejor explicación también soluciona el problema de Duhem. Para ambas cosas, dice Weber, hay que "mostrar que el experimento de Meselson-Stahl apoyaba la hipótesis semiconservadora sin la ayuda de pruebas adicionales para eliminar errores posibles en la interpretación de los datos (excepto la calibración de los instrumentos)." [*ibid.*, p. 38]. Es decir que el experimento de Meselsohn y Rolfe no era necesario para la confirmación.

Weber entiende aquí por explicación de un fenómeno "describir un mecanismo que produce el fenómeno" [*ibid.*, p. 33]. Y define mecanismo como entidades y actividades organizadas tales que producen cambios regulares desde las condiciones de comienzo a las de finalización [idem]. En el caso que nos ocupa, se supone que las moléculas de ADN tienen unos componentes y se comportan de una determinada manera. Lo que está en cuestión es el comportamiento del ADN en su replicación y se oponen dos mecanismos: el de la hipótesis semiconservadora y el de la hipótesis conservadora.

La importancia de los mecanismos depende de cómo Weber entiende la mejor explicación. La mejor explicación es la que proporciona mayor "familiaridad con las entidades y actividades, así como con ciertos patrones de dependencia contrafáctica involucrados en la producción del fenómeno explicado, en particular en tanto que instancia regularidades." [*ibid.*, p.36]. Por tanto, una buena explicación es "una descripción de un mecanismo, en otras palabras, una disposición de procesos causales entrelazados que juntos producen los hechos a explicar." [*ibid.*, p. 37].

Lo que Weber sugiere es que sin la necesidad del segundo experimento, el mecanismo semiconservador "[...] era suficiente para explicar los datos por sí mismo […] los mecanismos alternativos [la hipótesis conservadora] hubieran requerido la adición de mecanismos o "epiciclos" en orden a explicar los datos de Meselson y Stahl [en concreto, suponer que había moléculas pesadas y ligeras que estaban pegadas]. [...] Por el contrario, con el mecanismo semiconservador [con la apertura de la doble hélice] está absolutamente claro por qué es probable producir los patrones de bandas; nada es misterioso." [*ibid.*, p. 38]

De este modo, según Weber, la hipótesis semiconservadora es la que está respaldada por la mayor evidencia [*ibid.*, p. 46], y queda solucionado el problema de Duhem.

6. Pero al precio de olvidarse del apoyo empírico. Porque cabe pensar que lo que hicieron los científicos en el ejemplo de Weber fue descontar un resultado favorable. Y, en ese caso, el experimentalismo de Mayo lleva ventaja:

1º) Si el resultado favorable del segundo experimento no se hubiese dado, entonces Mayo podría presumir de que ella ya lo había advertido, dado que la contrastación no era severa.

2º) Aceptemos que ser la mejor explicación es una evidencia en favor de la hipótesis semiconservadora. Pero no permite confirmarla. Porque hay al menos dos mecanismos de replicación del ADN excluyentes entre sí. Y hasta el segundo experimento, el de Meselson y Rolfe, no supieron los científicos exactamente lo que el primer experimento, el de Meselson y Stahl, establecía. Sencillamente porque no se había establecido si lo que mostraban las bandas de densidad del primer experimento eran moléculas de ADN pegadas entre sí o cadena de ADN con nitrógeno pesado y ligero. Y ese defecto es precisamente lo que la severidad de Mayo nos señala.

3º) Como hemos visto, Weber critica a Mayo por su solución del problema de Duhem. Weber cree que conduce a un regreso infinito de contrastaciones. Y que eso se evita mediante la inferencia de la mejor explicación, pues ésta involucra los mecanismos explicativos que funcionan como un todo [*ibid.*, p. 39]. Pero Mayo aceptaría que al realizar una contrastación existe un conocimiento que no se pone en cuestión, y que se puede concretar en mecanismos o modelos [Mayo (1996), cap. 5]. En nuestro caso, hay un mecanismo subyacente *común* a las dos hipótesis que nos conciernen que no está puesto en cuestión y que, precisamente, permite eliminar el resto de hipótesis alternativas. Pero entre las dos hipótesis sólo se decide descartando empíricamente el supuesto auxiliar A' sobre el significado de las bandas de densidad. Y eso es lo que señala el requisito de severidad de Mayo.

Referencias bibliográficas

Mayo, D. (1996), *Error and the Growth of Experimental Knowledge*, University of Chicago Press.

— (2006), *Severe Testing, Error Statistics, and the Growth of Theoretical Knowledge*, *Borrador,* 25/9/2008, <http://www.error06.econ.vt.edu/>.

Weber,M. (2009), "The Crux of Crucial Experiments: Duhem's Problems and Inference to the Best Explanation", *British Journal for the Philosophy of Science* 60, pp. 19-49.

¿Puede otorgarse status teórico a una perspectiva de límite? Ciencias sociales y crítica poscolonial.

Rosa Sierra
Universidad de Frankfurt
unarosaesunarosa@yahoo.com

La perspectiva de límite (*border perspective*) es una idea expuesta por Walter Mignolo en sus trabajos sobre poscolonialismo, y se trata de una perspectiva *sui generis* que puede adoptarse en las ciencias sociales y las humanidades, y que integra aspectos como la *transculturalidad*, la *localización* y la *poscolonialidad* en la producción de conocimiento. La corriente poscolonialista en la que se inscribe el trabajo de Mignolo es descrita por él mismo como un discurso crítico de la modernidad [Mignolo (2000), p. 96], en el que se reconoce a la colonización de América como un elemento histórico-geográfico constitutivo de la misma. Con este reconocimiento tiene lugar, según Mignolo, una descentralización de la producción de conocimiento, y una relocalización de los focos posibles de enunciación del discurso científico. Esto hace que los discursos coloniales y poscoloniales no sean simplemente un tema de investigación disciplinar, sino una perspectiva para las ciencias sociales y humanas (Mignolo (1993), p. 134]. La reinterpretacion que hace Mignolo de la modernidad se concentra en el aspecto epistemológico de la experiencia colonial, y en las consecuencias de ella para la práctica científica y humanística. En este sentido su trabajo podría ser interesante para la filosofía de las ciencias sociales; sin embargo, luego de analizarlo puede verse que su relevancia es más bien política que estrictamente teórica.

La idea de Mignolo de una perspectiva *sui generis* a partir de la integración del aspecto de la colonialidad en la reflexión sobre la práctica científica y humanista está influida por su recepción del modelo de análisis histórico-económico propuesto por Immanuel Wallerstein para explicar el surgimiento de lo que él llama el sistema-mundo moderno [Wallerstein (2007)]. Entre las categorías básicas de este modelo se cuentan las nociones de *centro* y *periferia*, con las que Wallerstein explica los procesos económicos que tienen lugar dentro del sistema [Wallerstein (2007), p. 12 y (1982), pp. 92 y 99]. Mignolo adopta la idea del sistema-mundo moderno y la amplía a la noción de sistema-mundo moderno/colonial por medio de la integración de una tesis que toma de Enrique Dussel, según la cual, la colonización de América es un proceso constitutivo de la modernidad [Dussel (1993)]. Sin embargo, en lugar de hablar del centro y la periferia de este sistema-mundo moderno/colonial, Mignolo habla de *límites internos* y *externos* del sistema. Según él, las nociones de centro y periferia tienen un carácter eminentemente territorial o geográfico, y no harían justicia a los fenómenos que él analiza con ayuda del modelo. Dichas categorías son adecuadas para un análisis histórico-económico, pero en el marco de su crítica de la modernidad, a Mignolo le interesa integrar el aspecto cultural para formular una

tesis epistemológica. Aunque discutible, la idea de Mignolo es que dichas categorías no expresan la dimensión *simbólica* integrada en los fenómenos en cuestión, que es la que se necesita para obtener conclusiones relevantes para el nivel epistemológico.

La noción de límite interno/externo permite formular el punto de partida de su análisis: el lugar en que las teorías o los discursos son producidos, bien sea en los bordes internos o en los bordes externos del sistema, ha representado históricamente una diferencia para la validez de dicho conocimiento [Mignolo (2000), pp. 55-56]. Estos bordes, sin embargo, pueden constituir el foco mismo de una nueva perspectiva. Situados en el límite del sistema es que puede hacerse justicia a la transculturalidad, y desarrollar lo que Mignolo llama "una epistemología de y desde los bordes del sistema-mundo moderno/colonial" [*ibid*., p. 52]. Mignolo se concentra en el caso de los científicos, humanistas e intelectuales que están situados en el cruce de culturas, lo que en los estudios poscoloniales está relacionado con el tema de la hibridación cultural. Mignolo ilustra a través de ejemplos la perspectiva a la que se refiere citando la experiencia de sociólogos, historiadores y literatos –entre otros– que según él articulan la dimensión poscolonial en sus respectivos trabajos [*ibid*., p. 64-88]. La poscolonialidad de dichos discursos consiste en (i) sacar a la luz «el lado colonial del sistema-mundo moderno» y las relaciones de poder característicamente coloniales presentes en la modernidad. Haciendo esto, dichos discursos (ii) reordenan la geopolítica del conocimiento, es decir, «reubican la relación entre localización geohistórica y producción de conocimiento». [*Ibid*., p. 93].

La perspectiva de frontera es, entonces, la perspectiva epistemológica de un discurso crítico que trae a la luz las dimensiones de transculturalidad, localización y poscolonialidad. ¿Cuál es el elemento epistemológico de esta perspectiva? ¿En qué sentido representa una perspectiva *relevante* para las ciencias sociales?

Comencemos abordando la cuestión desde las ciencias sociales, considerando un modelo específico. En su *Teoría de la acción comunicativa*, Jürgen Habermas desarrolla un análisis de las diferentes perspectivas que pueden adoptarse en las ciencias sociales y distingue tres tipos: la perspectiva del *observador*, la perspectiva del *participante* y la perspectiva del *narrador* [Habermas (1995), Tomo I, pp. 164-174 y Tomo II, pp. 205-208]. Estas nociones habían sido desarrolladas en diferentes corrientes al interior de las ciencias sociales, y Habermas discute qué tan adecuada es cada una a la hora de explicar ciertos fenómenos sociales, en particular, aquellos que implican una situación comunicativa. También en el análisis dedicado a la fundamentación de su teoría social, Habermas mantiene el esquema de las perspectiva *interna* /*externa*, que de hecho es consecuentemente expresado en su concepto "doble" de sociedad como sistema y mundo de la vida (*Lebenswelt*), que según él tiene mayor alcance explicativo. [Habermas (1995), Tomo II, pp. 179, 227-228]

La primera serie de distinciones (observador, participante, etc.) está desarrollada en el nivel metodológico, y hace referencia al científico, y a la perspectiva que él toma en relación con un fenómeno que desea explicar. La otra

distinción (interna/externa) está inscrita en el nivel metateórico, conceptual, y hace referencia a la perspectiva que puede adoptarse en relación con el objeto de la explicación, dada su naturaleza[1]. Las distinciones son complementarias, de modo que la perspectiva del observador es una perspectiva externa al fenómeno y la perspectiva del participante es una perspectiva interna, así como la perspectiva del narrador, sólo que esta última contiene un elemento adicional –el elemento discursivo. Lo importante de tener presente que se trata de dos series de distinciones es que con eso se preserva la claridad respecto al nivel de análisis en el que nos estemos moviendo.

Con esta ilustración podemos seguir analizando los elementos que pone en juego la perspectiva de límite, y apreciar en qué nivel podemos situarla. Si tomamos en cuenta la noción de límite con la que está conectada la perspectiva (límite externo / interno del sistema mundo), entonces podemos considerarla en relación con el fenómeno que se estudia. En ese caso, las opciones que resultan son la de una perspectiva desde los límites internos del sistema y una perspectiva desde los límites externos; y la que Mignolo propone, que es una perspectiva desde *los límites mismos* [Mignolo (2000), pp. 52, 110]. En relación con el fenómeno que se explica o que se hace tema de un discurso, lo que esta perspectiva implica es que el fenómeno puede ser estudiado desde una localización distinta a la suya, o sin estar situado al exterior o al interior del sistema, pero considerando *ambas* localizaciones al mismo tiempo. Con esto se supera la restricción de producir el discurso desde una lógica exclusivamente, y se puede disponer de los recursos de dos lógicas distintas, sin sintetizarlas dialécticamente, y manteniéndolas en su irreductible diferencia [*ibid*., pp. 67, 85-86].

Tomemos en cuenta ahora los tres aspectos que la perspectiva de límite busca integrar: transculturalidad, localización y poscolonialidad. El primer rasgo puede entenderse como la integración de las diferentes tradiciones (tanto las culturales en sentido amplio, como también las académicas específicamente) que confluyen en la experiencia de un sujeto particular productor de un discurso. El rasgo de la localización expresa la existencia de un vínculo entre la producción de conocimiento y la localización historico-geográfica de esa producción. Finalmente, el rasgo de la poscolonialidad hace referencia a la conciencia de la colonialidad y a la consecuente subversión de ella (de su carácter subalterno) en la enunciación misma del discurso. De estos tres aspectos, el segundo se encuentra relacionado con la perspectiva en relación con el fenómeno, a la que me referí en el párrafo anterior. El primero y el tercero, en cambio, atañen directamente al *sujeto*, es decir, al científico o al productor de un discurso. En ellos se puede reconocer la dimensión metodológica de la propuesta de Mignolo: el procedimiento del investigador es el de situarse al borde de las diferentes culturas académicas que constituyen su experiencia, tomando una posición crítica respecto a todas ellas. Este es el procedimiento necesario exigido por la situación de la

[1] Según la teoría habermasiana, el análisis de una situación comunicativa exige una perspectiva interna a ella. El análisis de la sociedad exige, a su vez, ambas perspectivas: la interna y la externa [Habermas (1995), Tomo II, pp. 226-227].

«geopolítica del conocimiento»: en el sistema-mundo moderno/colonial, el conocimiento ha sido producido en el interior de los límites del sistema, y la producción por fuera de esos límites ha sido relegada a un status secundario o subalterno, y esta situación sólo puede superarse reconociendo los límites, situandose en ellos, y tomando en cuenta lo que hay a ambos lados del límite [*ibid*, pp. 67-68].

¿Cómo podemos responder a nuestra pregunta sobre el status teórico de la perspectiva de límite, teniendo en cuenta este panorama? Creo que es necesario que se cumplan (por lo menos) dos criterios. Por una parte, su plausibilidad epistemológica: la perspectiva es coherente con una cierta concepción del conocimiento; por otra parte, su relevancia explicativa: hay un fenómeno que precisa la adopción de esta perspectiva para poder ser explicado.

Respecto al primer criterio, la pregunta clave es la de la posibilidad de la transculturalidad en la experiencia epistémica. La adopción de la perspectiva presupone la idea de que el conocimiento no puede ser considerado independientemente de su génesis, del contexto de su enunciación, de las coordenadas histórico-geográficas y de las variantes culturales de ese contexto. Pero el reconocimiento de la localidad inherente al conocimiento no tiene en este contexto la forma de un relativismo cultural, que a continuación obligue a preguntar por la posibilidad de la comunicación intercultural. Lo que se presupone, en cambio, es que un sujeto puede tener acceso epistémico a su propia cultura y a otras diferentes, y puede articular discursivamente esa experiencia. El investigador no sólo puede (y debe) tomar en cuenta su propia procedencia cultural, específicamente en su formación académica, y ser crítico ante ella, y ante los elementos culturales que encuentre en la situación investigada; también puede hacer eso mismo con una cultura diferente de la suya. Por una parte, esta condición parece cumplirse empíricamente, por lo menos tal como lo presenta Mignolo: su elaboración de la perspectiva parte de la constatación de una serie de trabajos en sociología, historia y literatura por parte de académicos que se encuentra en un cruce de culturas, y que han integrado elementos pertenecientes a las diferentes tradiciones académicas y nacionales con las que han estado en contacto [*ibid*, pp. 64-88; en especial pp. 78, 82-83]. Por otra parte, los trabajos recientes en epistemología evolucionista, que atacan el relativismo cultural de la sociología del conocimiento y de las epistemologías postmodernas, ofrecen argumentos en favor de una perspectiva transcultural [Gontier *et.al* (2006), pp. 6-10].

Respecto al segundo criterio, la perspectiva de límite encuentra el siguiente contra-argumento: ¿qué hay en la «situación poscolonial» que no pueda ser explicado con otros modelos, por ejemplo, desde la sociología del conocimiento, o desde ciertas teorías sociales neo-marxistas que critican las relaciones económicas y sus consecuencias sociales, y el reflejo de estos factores en la esfera simbólica? Como fenómeno histórico-social, el colonialismo y los fenómenos asociados con la situación poscolonial (colonización interna, modernización, etc.) pueden ser investigados extensamente sobre la base de los análisis sobre el imperialismo o el totalitarismo, que de hecho son reconocidos como una de las fuentes teóricas de los discursos poscoloniales [Chrisman y Williams (1994), pp. 6-8]; o también

sobre la base de los análisis de los procesos de globalización, como lo hacen el mismo Wallerstein (2006), o incluso Dussel, entre otros [Jameson y Miyoshi (1998)].

Podemos concluir, entonces, que el status teórico de la perspectiva de límite es débil: sólo cumple el primer criterio que, aunque necesario, no es suficiente para sustentar la relevancia teórica o metodológica de la perspectiva. Sin embargo, aunque la perspectiva de límite carezca de una relevancia estrictamente teórica, si se la entiende bien (y espero que mi análisis haya logrado este objetivo), debe quedar claro la relevancia que ella posee en una dimensión política, específicamente por el elemento poscolonial en ella: en la medida en que, como discurso crítico, influye en la construcción de identidad en los territorios con un pasado colonial. Su contribución es de carácter performativo: la enunciación discursiva misma representa la subversión de la colonialidad, de la subalternidad que caracterizaba a quien está enunciando el discurso antes de dicha enunciación.

Referencias bibliográficas

Chrisman, L. y Williams, P. (eds.), (1994), *Colonial Discourse and Post-Colonial Theory*, New York, Columbia University Press.

Dussel, E. (1993), 'Eurocentrism and Modernity. Introduction to the Frankfurt Lectures', en Beverly, J. y Oviedo, J. (eds.), *The Postmodernism Debate in Latin America*, Durham, Duke University Press, pp. 65-76.

Gontier, N. *et al.* (2006), *Evolutionary Epistemology, Language and Culture*, Dordrecht, Springer.

Habermas, J. (1995), *Theorie des kommunikativen Handelns*, Frankfurt am Main, Suhrkamp, Tomos I y II.

Jameson, F. y Miyoshi, M. (1998), *The Cultures of Globalization*, Durham and London, Duke University Press.

Mignolo, W. (1993), 'Colonial and Postcolonial Discourse: Cultural Critique or Academic Colonialism?', *Latin American Research Review* 28 (3), pp. 120-134.

— (2000), *Local Histories / Global Designs*, Princeton, Princeton University Press.

Wallerstein, I. (1982), 'World-systems Analysis: Theoretical and Interpretative Issues', en Hopkins, T., Wallerstein, I. *et al.* (eds.), *World-systems Analysis: Theory and Methodology*, Beverly Hills, Sage Publications, pp. 91-103.

— (2006), *European Universalism*, New York and London, The New Press.

— (2007), *World-Systems Analysis: An Introduction*, Durham y London, Duke University Press.

Action-Reaction: Matter-Geometry interaction in GR

Adán Sus
Universidad Autónoma de Barcelona
adansus@gmail.com

In general relativity (GR) the so called response equations $T^{\mu\ \nu}{}_{;\nu} = 0$ are a direct consequence of Einstein's field equations. From them one can derive the general relativistic version of an energy-momentum conservation law and a geodesic principle to the effect that the world lines of test bodies are the geodesics of the spacetime metric. It is well known that this result can be seen also as a consequence of the covariance properties of the theory: applying Noether's second theorem to the coordinate independent gravitational Lagrangian and using satisfaction of the gravitational field equations one gets precisely the response equations. This result can be generalised for metric theories of gravity that are Lagrangian based; in a 1974 paper by Lee, Lightman and Ni (LLN), they prove that if all the fields appearing in the Lagrangian are dynamical (varied in the action), $T^{\mu\ \nu}{}_{;\nu} = 0$ is a consequence of the gravitational field equations.

Furthermore, there are other spacetime theories for which energy-momentum conservation cannot be seen as a consequence of the field equations but rather as a requirement that imposes certain restrictions on the covariance of the theories. Examples of this are unimodular relativity or the various field equations that Einstein tried before arriving at the final ones; as is known, at some point Einstein used energy-momentum conservation as a physical condition that would restrict the covariance of his sought-after field equations.

In this paper I explore whether this apparent difference between spacetime theories can help us in the search for a substantive notion of general covariance or a criterion that allow to distinguish GR from previous spacetime theories. An initial idea is that Noether's second theorem, through the LLN result, could provide a criterion to distinguish between formal and substantive versions of general covariance in the following sense: take a coordinate independent lagrangian spacetime theory and see whether the response equations are a consequence of field equations including fields other than the matter fields: if they are not, this is an indication of the presence of a non-dynamical spacetime structure. It turns out that this criterion is either still merely formal – if one understands that any condition obtained variationally can be considered a field equation – or ambiguous. To see this one can take, for instance, a generally covariant lagrangian version of a classical field theory in Minkowski spacetime, in which the condition of flatness ($R^{\mu\ \nu\ \lambda\ \rho} = 0$) is deduced using Hamilton´s principle thanks to the introduction of a lagrangian multiplier. What this type of example

shows us is that a criterion that uses a variational definition of being dynamical and does not distinguish gravitational field equations from coordinate conditions is going to be too weak. Nevertheless, the criterion can be modified by demanding that the field equations for which $T^{\mu\nu}$ act as a source alone be sufficient to derive $T^{\mu\nu}{}_{;\nu} = 0$. I discuss whether such a modification is enough to provide a substantive criterion. My view is that a criterion based on this idea serves not only to distinguish GR from theories with fixed spacetime structure but also from others in which only part of it is fixed independently from matter. This also shows an important feature of this approach in relation to others that seek to single out GR: the different status of the interrelation between spacetime and matter in spacetime theories is a question of degree.

I connect the search for this criterion with three interrelated conceptual issues arising from the interpretation of GR. First, this criterion can be seen as the realisation of an old metaphor about what distinguishes GR from previous spacetime theories, that in GR matter tells spacetime how to curve and spacetime tells matter how to move. In a more general fashion it could be formulated as an action-reaction approach to provide a notion of background independence: geodesic motion (at least for test bodies) is determined only through the field equations for which matter acts as a source. This approach – the idea that something like a generalised action-reaction principle between spacetime and matter is present in GR – has been suggested before [Anandan and Brown (1995)] but it has never been given a precise formulation (I do so in a longer version of this paper). A related attempt to embody the notion of action-reaction can be found in the absolute objects program of Anderson and Friedman; although, as I argue elsewhere, this program is effective in detecting objects that are not acted on but not so much when it comes to characterise the degree to which they can be said to act on matter.

The second conceptual connection is the following: this attempt at giving content to the uniqueness of GR amongst spacetime theories can be linked to one of the virtues that Einstein, at some points, attributed to GR: namely, that the theory, contrary to Newtonian physics and Special Relativity, provides a dynamical explanation of inertia. I discuss different senses in which one can say inertia to be explained dynamically and look at whether according to them this is a differential feature of GR. The first of these senses is closely connected to the idea of action-reaction discussed above. One can think that a spacetime theory explains inertia if the inertial structures of the spacetime theory in question – the ones that are responsible for inertial motion and inertial effects – are influenced by matter. In this sense it is clear that GR fares better than previous theories that introduce a fixed spacetime background, and it is this sense that is invoked by Einstein when attributing a defect to Newtonian mechanics. Nevertheless, from a conceptual point of view this approach is problematic on its own; the same idea of action-reaction suggests that one is postulating some type of physical interaction between spacetime and matter or a kind of causal relationship. And while Einstein saw this as a conceptual fault of Newtonian physics, it is also true that even having a

dynamical spacetime, the suggestion of spacetime causing inertial motion and inertial effects should come along with some type of causal story to support it; how are bodies supposed to feel the presence of a certain spacetime? The lack of hope in being able to answer this question, and its dubious status as a legitimate question, makes this sense of explaining inertia conceptually unsatisfactory.

One can try a different strategy in order to understand a theory as providing a dynamical explanation of inertia, one related to what have been termed constructivist approaches to spacetime. The leading idea here is that the physical laws – by this I mean the field equations and equations of motion – on top of containing information about the behaviour of fields and bodies, also encode the spacetime structures and in particular the ones responsible for inertial motion. Thus, the explanatory weight falls on the laws themselves rather than on the spacetime structures, avoiding the problem of reification of such entities. The particular form that this approach takes will depend on what property of the equations one takes as primitive and what laws one considers as fundamental but in general, once these decisions are taken, one will be able to decide whether in one theory inertia is better explained than in another one. Nevertheless, this also contains the potential weakness of this approach; how to justify certain common features of laws, beyond them being brute facts, and how to discriminate between physical laws in relation to their explanatory power.

A third approach to the question about the status of inertia in GR comes from considering the role of the equivalence principle in the theory. It is well known that there are different versions of such a principle but most authors distinguish between a weak equivalence principle (WEP), asserting the equality between inertial and passive gravitational mass, and Einstein's equivalence principle (EEP), that expresses the physical equivalence between a freely falling system in a uniform gravitational field and an inertial system. There is, nonetheless, what I believe to be a stronger version of the equivalence principle, one expressed in Einstein's statement about identity of inertia and gravity. I propose to give a more precise content to this principle in order to understand the conceptual mechanism that GR employs to produce an explanation of inertia. This strategy involves thinking that inertia gets explained by it being reduced to, or equated with, a physical entity that we consider well understood – the gravitational field – for which we have dynamical laws and a connection to matter. So this active reading of the equivalence principle expressed by the complete equivalence between inertia and the gravitational field would be more explanatory than EEP in the following sense: while EEP tracks down inertia and gravity to a common geometrical origin but leaves open the possibility of partial determination of the geometry independently from matter, the active reading forbids this possibility of independent from matter determination. Hence the relation between this way of understanding the equivalence principle and the action-reaction principle. I defend that the equivalence principle provides a good avenue to understand why one can say that in GR inertia receives a dynamical explanation, and that it does so because it has an action-reaction form. Nevertheless, by using the equivalence

principle one borrows the explanatory import of the action-reaction principle while avoiding its causal flavour.

The third conceptual issue related to the idea of action-reaction is the question about the Machian character of GR, and this will be developed elsewhere.

References

Anandan, J. and Brown, H. R. (1995) "On the Reality of Space-Time Geometry and the Wavefunction", *Foundations of Physics* 25, pp. 349-360.

Lee, D. L., Lightman, A. P. and Ni, W. T. (1974) "Conservation laws and variational principles in metric theories of gravity", *Phys. Rev.*, D10, 6, p. 1685.

Cómo defender el realismo en economía

Obdulia Torres González
Universidad de Salamanca
omtorres@usal.es

Abordamos las disputas acerca del realismo científico desde cuatro ámbitos diferenciados: uno ontológico, donde nos preguntamos acerca de la existencia de las entidades inobservables postuladas por nuestras teorías; uno semántico donde, desde una perspectiva realista, afirmaríamos la verdad o la aproximación a ésta de las afirmaciones teóricas; uno epistemológico, donde nos interrogamos acerca de la cognoscibilidad del mundo y su independencia respecto al sujeto cognoscente y finalmente, cuestionamos cuáles han de ser los fines de la ciencia, ¿ésta progresa teniendo como meta proporcionar explicaciones verdaderas, o aproximadamente verdaderas, acerca del mundo o sólo debe proporcionarnos predicciones adecuadas que nos permitan el control y la manipulación del mundo?

Cada uno de estos ámbitos presenta múltiples matices de los que no vamos a hablar aquí.[1] Lo que nos interesa es una pregunta formulada por I. Hacking acerca de la posibilidad de ser realista en física y antirrealista en ciencias sociales. [Hacking, (1996) p. 43). Si coherentemente respondemos que no, la cuestión ineludible es cómo defender el realismo en ciencias sociales en general, y en la economía en particular, y qué tipo de realismo es posible defender. El problema radica en que gran parte de los argumentos esgrimidos a favor del realismo en física parecen ser de escasa aplicabilidad en el terreno de la economía.

La defensa que el propio Hacking hace del realismo científico, condensada en la ya famosa frase "hasta donde a mi concierne, si se puede rociar algo con ellos, entonces son reales" [Hacking, (1996) p.41] parece tener poca aplicación en el campo económico en la medida en que las relaciones causales que somos capaces de establecer en economía son bastante más difusas de las que es posible establecer en física y, especialmente, las dificultades de experimentación asociadas a la economía.

Otros argumentos como el "argumento del milagro" formulado inicialmente por Putnan (Putnan, 1978) no parece tener mejor fortuna. En este caso se hace depender la existencia de los términos teóricos del éxito predictivo de la teoría. El problema es que el éxito predictivo es bastante escaso en la disciplina económica, al menos para hacer descansar en él una posición realista.

Otros como Boyd (Boyd, 1973) defienden el realismo como una hipótesis abierta a comprobación empírica a través de la evidencia histórica. Pero tampoco parece que podamos afirmar el éxito empírico de las teorías económicas a través del decurso del tiempo.

[1] Para una caracterización exhaustiva del realismo científico véase González (1993) ó Diéguez (2005).

¿Qué es lo que caracteriza a la Economía que hace que las defensas tradicionales del realismo científico sean inaplicables?

U. Mäki [Mäki, (1996)] señala hasta cuatro diferencias entre la economía y la física que hace que estas defensas sean inaplicables. Estas serían: la ausencia de un uso filosófico del términos realismo entre los economistas, la ontología económica no se acomoda a la existencia independiente de la mente típica de las formulaciones del realismo científico, la cuestión de los términos teóricos y la imposibilidad de establecer la existencia y verdad en economía invocando la manipulabilidad o el éxito. Nos ceñiremos a la cuestión de los términos teóricos.

Tanto U. Mäki como D. Hausman establecen una diferencia clara entre la economía y la física en lo que respecta al problema de los términos teóricos. Y ambos introducen un tipo adicional de realismo que es el realismo de sentido común.

Hausman [Hausman, (1998)] plantea que aunque parte de los términos postulados por la economía son inobservables, a diferencia de la física, la economía no plantea nuevos inobservables. ¿Qué quiere decir esto? Pues que los inobservables propuestos por la economía han sido parte de la comprensión del sentido común del mundo durante milenios. Si nos ceñimos al ámbito de la teoría microeconómica tenemos que la teoría de la acción racional, que explica la conducta económica en base a preferencias y probabilidades subjetivas, proviene de la *folk psicology* que explica las acciones en base a creencias y deseos. La conclusión de Hausman es que no podemos ser antirrealistas acerca de las entidades (preferencias y probabilidades subjetivas) postuladas por la teoría económica y realistas acerca de la comprensión cotidiana del mundo.

La primera objeción a la postura de Hausman es obvia. No es lo mismo postular que un sujeto tiene una ordenación de preferencias reflexiva, transitiva y completa y que además el individuo maximiza una función de utilidad que afirmar que nuestro hijo prefiere el helado de chocolate al de vainilla, o que las probabilidades subjetivas, tal como son postuladas por la teoría de Bayes o por la cardinalización de la utilidad propuesta por V, Neumann y Morgenstern, es lo mismo que la creencia en que los aviones vuelan, según los ejemplos propuestos por el autor. La respuesta de Hausman es que la ordenación de preferencias y las probabilidades subjetivas son variaciones idealizadas de las nociones de deseos y creencias. En defensa de esta postura plantea tres argumentos:

- el papel funcional de ordenación de preferencias y de probabilidades subjetivas es el mismo que el de deseos y creencias en lo que se refiere a explicación y predicción de la acción.
- Al comprobar y operacionalizar la teoría de la elección racional los teóricos confían en la asociación entre la ordenación de preferencias y la fuerza de los deseos, y entre las probabilidades subjetivas y el grado de creencia de un individuo.
- La plausibilidad de los axiomas de la teoría de la elección racional depende de la asociación entre ordenación de preferencias y deseos y entre probabilidades subjetivas y grados de creencia.

Es decir, Hausman fundamenta la idealización en valores como la funcionalidad, la confianza o la plausibilidad. La cuestión es si el ámbito valorativo, aunque de valores epistémicos se trate, nos sirve para fundamentar la esfera ontológica y en este sentido nuestra primera aproximación es que, como mínimo, es una fundamentación controvertida.

En cuanto a U. Mäki para poder dar cuenta de su postura tenemos que hacer referencia a su taxonomía del realismo. Mäki distingue entre un realismo ontológico que afirma "x existe"; un realismo referencial que sostiene que las expresiones lingüísticas, pueden, deben o de hecho refieren a entidades del mundo real; un realismo representacional donde se postula no sólo que las expresiones lingüísticas refieren, sino que se le atribuyen propiedades a esas entidades y finalmente, un realismo veritativo que afirma que la verdad y la falsedad están entre las propiedades semánticas posibles de teorías y proposiciones. [Mäki, (1992), pp. 173-174]

Esta taxonomía ha sido objeto de críticas dado que es muy poco restrictiva y permite combinarlas con posturas tradicionalmente consideradas antirrealistas [Boyland & O'Gorman (1995) pp. 118-122]. El punto que nos interesa bisagra entre la definición de realismo representacional y veritativo. Según la taxonomía propuesta por Mäki, podemos ser considerados realistas representacionales aunque las propiedades que atribuyamos a las entidadades sean falsas. Pongamos por ejemplo el caso de las preferencias de los agentes económicos. La teoría macroeconómica atribuye a esas entidades las propiedades de ser reflexivas, transitivas y completas. Ya en virtud de esa atribución seremos realistas representacionales, dado que no es necesario que esas propiedades sean verdaderas, que efectivamente se den en el mundo, sencillamente atribuimos propiedades a nuestras entidades. El realismo veritativo, aunque más restrictivo, no acude en nuestra ayuda, pues sólo afirma que la verdad o la falsedad está entre las propiedades semánticas posibles de nuestras proposiciones. Es decir, la proposición "las preferencias de los agentes son reflexivas, transitivas y completas" es susceptible de ser verdadera o falsa. Ello posibilita que un autor como M. Friedman, considerado como un estereotipo del instrumentalismo, pueda ser considerado un realista en cualquiera de los sentidos propuestos por Mäki. Es obvio que lo que está en juego es la teoría del significado que se defienda, la cuestión es si los términos que proponen nuestras teorías refieren si no son verdad de nada en el mundo real. Por razones de espacio no podemos entrar aquí en este debate pero nuestra intuición estaría más próxima a la teoría de Frege que a la de Putnam. En cualquier caso la postura de Mäki no deja de ser un tanto paradójica dado que sostiene que no es la existencia de los constituyentes básicos de la teoría económica la principal cuestión del realismo en economía, sino la cuestión de la verdad. [Mäki (1998), p. 309)

Mäki sostiene que el problema del realismo en economía no toma la forma del problema de los términos teóricos, como en la física, entidades inobservables de los cuales se afirma la existencia. "En el caso de fotones y quarks hay una salida ontológica de la esfera del sentido común de mesas y árboles: hay una diferencia de clase entre fotones y quarks de un lado y mesas y árboles de otro. En el caso de

preferencias y expectativas, empresas y beneficios, familias y salarios, no hay tal salida: no hay ninguna diferencia de clase. El mobiliario ontológico de la "economía *folk*" es compartida por la economía científica, mientras el mobiliario ontológico de la "física *folk*" es reemplazada por la "física científica". La economía no parece emplear el tipo de conceptos teóricos que encontramos en la física." (Mäki, 1998, 307)

¿A que se refieren tanto Mäki como Hausman cuando hablan de la esfera del sentido común? Téngase en cuenta que Mäki afirma "en lo que concierne a los referentes individuales de la teoría el realismo científico ontológico podría adecuarse a la física, mientras la economía parece requerir algún tipo de realismo ontológico de sentido común. ¿En qué consiste ese realismo? Hausman no lo aclara mientras que Mäki proporciona algunas pistas en su texto de 1992.

Mäki relaciona el sentido común con las capacidades cognitivas y conceptuales de las personas legas en contraste con los especialistas en la disciplina en cuestión. Así define un realismo de sentido común mínimo como realismo acerca de objetos y representaciones del sentido común, y un realismo de sentido común radical que añade antirrealismo acerca de los objetos y representaciones y científicas. Lo que no nos aclara es en qué basa el lego, como opuesto al experto, su creencia en la existencia de los objetos de sentido común. ¿En un realismo ingenuo, en el sentido de Bunge, que sostiene que el mundo es lo que aparenta ser o un realismo crítico que afirma que los sentidos pueden engañarnos y necesitamos hacer uso de la razón para representar tanto lo perceptible como lo imperceptible? [Bunge (1985) pp.16-17] Una afirmación realizada más abajo parece apuntar a lo primero: "En la medida en que un realista de sentido común es incapaz de hacer una distinción seria entre la forma en que el mundo es y la forma en que el mundo aparece a la observación…" Pero tal y como sostiene Bunge "ésta es la gnoseología de los niños de corta edad" [Bunge (1985) p.16] Sin embargo es este tipo de realismo el que el autor propone como necesario en Economía, vale la pena citarlo en extenso.

"Desde el punto de vista del realismo radical del sentido común, la existencia de muchas de las entidades fundamentales de las teorías económicas no parece dudosa, mientras la existencia de muchas entidades físicas parecen sospechosas. La diferencia aparece al nivel de la referencia: es relativamente más fácil aceptar las teorías económicas como factualmente referenciales de lo que es el caso con las teorías físicas. Ningún cambio ocurre a este respecto en la física cuando nos movemos de referencia a representación. Las propiedades atribuidas a las entidades postuladas por las representaciones teóricas son a menudo tan dudosas como aquellas entidades mismas a los ojos de un realista radical del sentido común. La situación es diferente en economía. Aquí a las entidades a las que se refiere y que se cree que existen sobre la base de la experiencia cotidiana se le han atribuido propiedades (tales como maximización racional, información perfecta, homogeneidad o divisibilidad perfecta) que no se cree que poseen, estando esta creencia basada en la observación ordinaria." [Mäki, (1992) p. 182]

En primer lugar con este realismo ingenuo es dudoso que las entidades fundamentales de la teoría económica no parezcan sospechosas, dado que son