

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FILOSOFÍA
DEPARTAMENTO DE LÓGICA Y FILOSOFÍA DE LA CIENCIA



EXTERNISMO Y ACCESO PRIVILEGIADO

TESIS DOCTORAL DE:

SILVIA ANDRÉS BALSERA

DIRIGIDA POR:

LUIS FERNÁNDEZ MORENO

Madrid, 2013

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Filosofía

Departamento de Lógica y Filosofía de la Ciencia



EXTERNISMO Y ACCESO PRIVILEGIADO

Memoria para optar al grado de doctor presentada por

Silvia Andrés Balsera

Bajo la dirección del doctor Luis Fernández Moreno

Madrid, 2012

Índice

	Página
Agradecimientos	v
Introducción	1
Capítulo 1: Externismo semántico y externismo psicológico	15
1. Externismo semántico (I): La Teoría Clásica y la Teoría de la Referencia Directa	17
1.1 Los términos singulares y sus propiedades semánticas	17
1.2 La Teoría Clásica de Frege sobre el valor semántico de los nombres propios	20
1.3 La Teoría Clásica de Russell: descripciones, nombres lógicamente propios y distinción entre conocimiento por familiaridad y conocimiento por descripción	25
1.4 La Teoría de la Referencia Directa	31
2. Externismo semántico (II): Millianismo, Kripke y Putnam	36
2.1 Las críticas de Kripke a la Teoría Clásica de los nombres propios y de los términos de género natural. Sus propuestas al respecto	37
2.2 Las tesis de Putnam sobre la semántica de los términos de género natural y los experimentos mentales de la Tierra Gemela	48
3. Externismo psicológico (I): uso de los experimentos de la Tierra Gemela en Filosofía de la Mente	61
4. Externismo psicológico (II): el anti-individualismo de Burge	69
5. Conclusiones	76

Capítulo 2: El conflicto <i>prima facie</i> entre el externismo y el acceso privilegiado y el trilema de Bilgrami	79
1. Las líneas fundamentales del conflicto <i>prima facie</i>	81
2. La propuesta compatibilista de Burge: el modelo basado en la inclusión	89
3. Donald Davidson: el externismo en el marco de la interpretación radical	99
4. El trilema de Bilgrami: externismo ortodoxo y atribuciones inconsistentes	112
5. Conclusiones	128
Capítulo 3: Los casos de traslado	131
1. Casos de traslado y memoria	133
2. Casos de traslado y razonamiento	153
3. La relevancia de los casos de traslado	182
4. Conclusiones	188
Capítulo 4: Externismo y acceso <i>a priori</i> al mundo externo	193
1. Versiones del argumento incompatibilista de la consecuencia	194
1.1 La paradoja de McKinsey	195
1.2 La <i>reductio</i> de Boghossian contra el externismo natural	199
1.3 La <i>reductio</i> de Brown contra el externismo social	207
2. Respuestas compatibilistas a las distintas versiones del argumento de la consecuencia	210
2.1 Cuestionamiento de las premisas del argumento de la consecuencia	211
2.1.1 Contra el carácter <i>a priori</i> de las premisas del argumento de la consecuencia	211

2.1.2 Cuestionamiento de la verdad de las premisas del argumento de la consecuencia	214
2.1.2.1 La verdad de las premisas de la paradoja de McKinsey	215
2.1.2.2 La verdad de las premisas de la <i>reductio</i> de Boghossian	220
2.1.2.3 La verdad de las premisas de la <i>reductio</i> de Brown	225
2.2 Cuestionamiento del carácter paradójico de la conclusión	229
2.2.1 La hipótesis escéptica de los cerebros en cubetas	230
2.2.2 Los argumentos de Putnam contra la hipótesis escéptica de los cerebros en cubetas	233
2.2.3 La posibilidad de un uso anti-escéptico del argumento de la consecuencia	238
2.3 Petición de principio y transmisión de la garantía	244
2.3.1 Consideraciones de Wright sobre el fallo en la transmisión de la garantía	246
2.3.2 Consideraciones de Davies sobre el fallo en la transmisión de la garantía	261
3. Conclusiones	271
Apéndice: Las propuestas de McKinsey posteriores a (1991)	274
Capítulo 5: Hacia una concepción viable de la individuación externalista de los contenidos mentales	283
1. Distintas concepciones del internismo/externismo psicológico	285
2. Problemas de la clasificación usual de las variedades del externismo con respecto a la individuación	294
2.1 Tipología usual de las variedades de externismo con respecto a la individuación de contenidos mentales	294
2.2 Fenómenos psico-lingüísticos que respaldan el externismo con respecto a la individuación	301
2.3 Tensiones internas a la clasificación usual de las variedades del externismo	308

3. Propuesta de revisión de la tipología usual del externismo	319
4. Por un externismo sin esencias	324
4.1 La relevancia del escenario de la Tierra Gemela	326
4.2 Cuestionamiento del caso particular del agua	328
4.3 La viabilidad de la concepción de los géneros naturales del externismo estrictamente natural	331
4.4 Influencia natural directa y atribuciones mentales. Nuestra propuesta: influencia natural mediada no-esencialista	348
5. Conclusiones	361
Capítulo 6: Conclusiones finales	365
Bibliografía	379

Agradecimientos

La elaboración de una tesis doctoral es un trabajo duro, lento y un tanto solitario. Durante estos últimos tres años han sido muchas las personas que me ayudaron a que lo fuera un poco menos. Valgan estas líneas para expresar mi agradecimiento a todas ellas (y perdón por los olvidos).

En primer lugar, debo decir que no habría podido realizar este trabajo sin el apoyo económico del programa de Formación del Profesorado Universitario, dependiente del Ministerio de Educación, Cultura y Deporte, así como del proyecto de investigación “Contribuciones al debate entre externismo e internismo” (referencia FFI2008-03092), financiado por el Ministerio de Ciencia e Innovación.

Sin duda, Luis Fernández Moreno ha sido el mayor apoyo intelectual que he tenido durante estos años y, si hay algún defecto en esta tesis, en ningún caso podrá serle imputable. Asimismo, debo agradecer su labor a otros profesores del departamento de Lógica y Filosofía de la Ciencia (especialmente a Antonio Blanco Salgueiro, Javier Vilanova Arias y Antonio Benítez López), ya que mi interés en los problemas filosóficos relacionados con el lenguaje y con la mente ha sido, en gran parte, fruto de haber asistido a sus clases así como a las de mi director de tesis. No quiero olvidarme de agradecer al personal de la Biblioteca su excelente trabajo y, especialmente, que lo desempeñen con tanta amabilidad y disposición.

Reservo una última mención para aquellos que me han brindado el mejor soporte personal que quepa imaginar, es decir, mi familia y mis amigos, tanto aquellos que son, además, compañeros de doctorado (Clara, Iván, Belén, Ana, Hernán) como aquellos que llevan conmigo desde hace mucho más tiempo (Mariluz, Pilar, Eva, Pablo). Y a Noelia, como siempre, gracias por todo.

Introducción

En nuestra vida cotidiana solemos explicar nuestras acciones y, muy especialmente, las acciones de los demás apelando a sus creencias, deseos, temores y esperanzas. Si, por ejemplo, María expresa frecuentemente sus dudas sobre la idoneidad de que cierto partido político en la oposición tome las riendas del gobierno, es natural atribuirle dicha creencia y una serie de deseos y temores relacionados con ella. Así, no parece verosímil que desee un triunfo de dicho partido en las próximas elecciones, sino que más bien lo teme. Si, además, la conocemos bien y sabemos en qué sentido van encaminadas sus expectativas, diremos que María prefiere que vuelva a ganar los comicios el partido que actualmente está en el gobierno y, consiguientemente, esperaremos que su propio voto vaya orientado en ese sentido. Estas creencias, deseos y temores que atribuimos a María tienen como fin volver inteligible para nosotros su comportamiento e integrar sus acciones puntuales en una cierta unidad dotada de coherencia. Inferimos, pues, qué creencias o deseos posee un hablante a partir de lo que hace y esperamos que esos estados mentales así inferidos nos orienten sobre su quehacer futuro. Esta consideración de los agentes, por la cual los tomamos como poseedores de creencias, deseos, temores, esperanzas, etc., que explican su comportamiento, forma parte de lo que se ha dado en llamar *psicología popular* o *psicología de la creencia y el deseo*. Esta última denominación se debe a que son esos dos estados mentales, la creencia y el deseo, los que cobran mayor protagonismo en nuestros informes psicológicos cotidianos.

Desde la Filosofía se han formulado propuestas muy distintas para examinar estas prácticas, e incluso hay quien no duda en otorgarles el estatus de una auténtica teoría.¹ Sin embargo, antes de abordar cualquier cuestión sobre el fenómeno mencionado, es preciso consensuar el uso de cierta terminología técnica. Para empezar, convengamos en denominar *pensamientos* o *estados mentales de carácter proposicional* a esta clase de estados mentales que atribuimos a los sujetos para explicar su comportamiento.² Emplearemos así el término “pensamiento” como una etiqueta genérica para hablar de las creencias, dudas, deseos, temores, etc., que *atribuimos* o *adscribimos* a los agentes. El hecho de que también usemos la denominación “estado mental de carácter

¹ Cf. Fodor (1987), pp. 6 y ss. y Jackson/Pettit (1990), p. 33.

² Aunque en la literatura sobre esta cuestión a menudo las expresiones “actitud proposicional” y “pensamiento” se usan de manera intercambiable, preferimos usar como expresiones equivalentes “pensamiento” y “estado mental de carácter proposicional”, reservando el término “actitud” para designar a uno de los componentes de los pensamientos, como veremos más adelante.

proposicional” para referirnos a los pensamientos nos pone en la pista para descubrir cuál es el común denominador de dichos estados, a saber, su dirección hacia un *contenido de carácter proposicional*. Para explicar esto, tomemos como ejemplo dos oraciones que nos informan sobre los pensamientos de dos sujetos, María y Luis:

(1) María cree que el partido en la oposición no es el apropiado para dirigir el país.

(2) Luis desea que el partido en la oposición gobierne en la próxima legislatura.

Las oraciones (1) y (2) constituyen informes de los pensamientos de María y de Luis, respectivamente. El contenido del pensamiento de María del que informamos en (1) es *que el partido en la oposición no es el apropiado para dirigir el país*. Podemos, por tanto, especificar el contenido de un pensamiento mediante lo que denominaremos la *cláusula-que*, es decir, la oración que figura a continuación del nexos “que” en el informe. Según esta pauta, el contenido del pensamiento de Luis en el informe (2) es *que el partido en la oposición gobierne en la próxima legislatura*. Se puede apreciar, asimismo, que los pensamientos atribuidos a estos hablantes en los informes (1) y (2) no sólo tienen un contenido distinto, sino que ambos muestran *actitudes psicológicas o mentales* diferentes hacia dichos contenidos, a saber, la actitud de creencia y la actitud de deseo, respectivamente. Los pensamientos o estados mentales de carácter proposicional son, por todo ello, estados mentales cuya identidad viene determinada (i) por el tipo de actitud psicológica (creencia, deseo, esperanza, etc.) que el hablante adopta hacia el contenido del pensamiento y (ii) por el propio contenido, especificado en la cláusula-que de los informes.³ De este modo, los informes (1) y (2) atribuyen distintos pensamientos a María y a Luis en cuanto que el contenido de los mismos es diferente y, además, la actitud mental con la que un hablante y otro se dirige hacia tales contenidos también es distinta.

Sin embargo, es necesario llevar a cabo una precisión adicional. Las cláusulas-que son oraciones que *capturan* o *especifican* el contenido de los pensamientos de los hablantes, pero este contenido mental no es idéntico a la oración que lo captura, sino que tiene una naturaleza diferente. Para esclarecer esto, necesitamos realizar una serie de distinciones previas. Por un lado, el término “oración” denota aquellas expresiones lingüísticas mediante las cuales podemos realizar actos de habla. Ahora bien, cabe distinguir entre *oraciones-ejemplar* y *oraciones-tipo*. Las primeras son el resultado de

³ En lo sucesivo, cuando empleemos el término “contenido” sin adjetivar nos estaremos refiriendo al contenido proposicional de nuestros pensamientos, al cual denominaremos también “contenido *mental*”, “contenido *psicológico*” o “contenido de nuestros pensamientos”.

las preferencias de los hablantes, incluyendo entre ellas no sólo emisiones orales, sino también inscripciones. Mientras que las oraciones-ejemplar son entidades localizadas en el espacio y en el tiempo, las oraciones-tipo, que las oraciones-ejemplar instancian, tienen un carácter abstracto. Cuando Juan y María se declaran mutuamente su aversión profiriendo la oración “Te odio”, cada una de sus emisiones lingüísticas es una ejemplificación o instancia particular de la oración-tipo “Te odio”. Por lo que respecta a las proposiciones, suscribiremos en lo que sigue la caracterización de las mismas ofrecida por Burge (2007, p. 2), en la medida en que su propuesta recoge varios rasgos de las proposiciones acerca de los cuales existe un notable consenso. De acuerdo con este autor, las *proposiciones* son entidades articuladas, portadoras primarias de valores de verdad, que tienen un carácter abstracto y que, además, son expresadas por las oraciones emitidas por los hablantes. El carácter articulado de una proposición tiene que ver con el hecho de que usualmente podemos distinguir varias partes dentro de las mismas, mientras que su carácter abstracto acerca estas entidades a las oraciones-tipo. Sin embargo, estas dos clases de entidades no deben confundirse. Dos oraciones-tipo tan diferentes como “El chino es un idioma difícil” y “Chinese is a difficult language” expresan lo mismo, a saber, la proposición de que el chino es un idioma difícil. Por último, al considerar que las proposiciones son los portadores primarios de valores de verdad queremos decir que estas entidades reciben en primera instancia el calificativo de verdaderas (o falsas) y, si una oración expresa una proposición verdadera (o falsa), entonces la oración también será verdadera (o falsa), en función de la verdad (o falsedad) de la proposición que expresa.

Podemos ofrecer, a partir de esto, una caracterización general de la forma que tienen los *informes de pensamientos* o, como también los llamaremos, *adscripciones psicológicas o mentales*:

S piensa que *P*

“S” ocupa el lugar de la expresión que designa al sujeto al que adscribimos el pensamiento en cuestión, mientras que “piensa” es la denominación general de las distintas actitudes mentales (como creer, desear, dudar, esperar, etc.) que el hablante puede tomar hacia un contenido. Por último, “P” abrevia aquella oración asertiva que especifica o captura el contenido hacia el cual se dirigen los pensamientos de S.

No obstante, conviene realizar ciertas puntualizaciones sobre el criterio de identidad para los pensamientos que hemos ofrecido. En este sentido, diremos que la identidad de los contenidos de nuestros pensamientos viene determinada, a su vez, por

la identidad de sus *conceptos* componentes.⁴ A fin de aportar una visión poco comprometida de lo que sean los conceptos, pues se trata de una cuestión sumamente discutida, diremos que por “concepto” debemos entender la contrapartida mental de los términos lingüísticos, bien singulares o bien generales. Cuando atribuimos a María su deseo de que el partido que gobierna en la actualidad salga triunfador en las próximas elecciones parece natural considerar que este hablante posee un cierto dominio sobre los conceptos expresados por los términos presentes en nuestro informe (“partido”, “oposición”, “dirigir”, “país”). Ese dominio tiene que ver con su capacidad para aplicar con cierta corrección esas expresiones y con el hecho de que este agente posee un número indeterminado de creencias verdaderas expresadas por oraciones que contienen dichos términos. La atribución de un concepto presupone, por tanto, ciertas habilidades en el agente a quien se le atribuye.

Otra distinción que resulta conveniente presentar en esta introducción es la diferencia entre las interpretaciones *de dicto* y *de re* de las atribuciones de pensamientos, así como la dicotomía *extensión-intensión*, debido a que es frecuente encontrar una caracterización de los verbos que rigen la cláusula-que como verbos que generan *contextos intensionales*. Con el fin de entender qué significa esto y su vinculación con las interpretaciones *de re* y *de dicto* de nuestros pensamientos, es necesario examinar, primero, el par extensión-intensión, para lo cual acudiremos a las consideraciones de Carnap, por ser el autor que introdujo el uso de esta dicotomía en la Filosofía del Lenguaje contemporánea. Por una parte, consideramos que la extensión de un término general es el conjunto de individuos designado por el término (Carnap 1956, p. 18).⁵ La extensión del término “oso” sería el conjunto de los osos.⁶ Por otra parte, la extensión de un término singular es el individuo designado por dicho término. Juan Manuel Navarro Cordón constituye, así, la extensión del término “el decano de la Facultad de

⁴ De esta forma, la identidad de un pensamiento depende del tipo de actitud psicológica que un hablante adopte hacia el contenido y del contenido proposicional mismo. Por su parte, la identidad del contenido del pensamiento dependerá de la identidad de los conceptos que involucra.

⁵ Preferimos la denominación “términos generales” y “términos singulares” para hablar de la distinción que Carnap establece entre predicados y expresiones singulares. Consideramos que la función de los términos generales es aplicarse a un conjunto de entidades, mientras que la de los términos singulares es designar un único individuo. No pensamos que este cambio en la terminología tenga demasiada importancia dado que aquello que nos interesa es su examen de la dicotomía extensión-intensión y no los detalles de la propuesta semántica carnapiana.

⁶ Otros autores prefieren decir, más bien, que la extensión de un término general comprende todos y cada uno de los individuos o de las entidades a las que se aplica el término; así, el término “oso” tendría como extensión a todos y cada uno de los osos (Cf. Devitt/Sterelny 1999, p. 88). Sin embargo, nosotros tomaremos en consideración únicamente la concepción de la extensión de un término general que hemos presentado en el cuerpo del texto, a saber, el conjunto de individuos a los que se aplica el término.

Filosofía de la Universidad Complutense en 2009”.⁷ Por último, la extensión de una oración es su valor de verdad (verdadera o falsa). En cuanto a la intensión de un término general, se trata de la propiedad que el término expresa, mientras que la intensión de un término singular es “el concepto individual expresado por él” (Carnap 1975, p. 41).⁸ La intensión de una oración es, por su parte, la proposición que esta oración expresa.

Visto esto, podemos establecer las condiciones para la *equivalencia extensional e intensional* de las expresiones. Diremos, pues, que dos términos generales son extensionalmente equivalentes si y sólo si designan el mismo conjunto de individuos, mientras que dos términos singulares que designan el mismo individuo y dos oraciones con el mismo valor de verdad son, respectivamente, términos singulares y oraciones extensionalmente equivalentes. Por el contrario, dos términos generales son intensionalmente equivalentes si y sólo si expresan la misma propiedad y dos términos singulares lo son si y sólo si expresan el mismo concepto individual. Dos oraciones son intensionalmente equivalentes si y sólo si expresan la misma proposición.

A partir del análisis de la extensión y la intensión de nuestras expresiones lingüísticas podemos diferenciar dos posibles interpretaciones de las atribuciones de pensamientos: una interpretación extensional o *de re* y una interpretación intensional o *de dicto*. Esta distinción tiene que ver con el modo en que tomamos en consideración la relación entre un sujeto, los contenidos (intensionales) de sus pensamientos y los objetos y estados de cosas (o extensiones) representados por dichos contenidos. Cuando interpretamos *de re* una atribución tomamos el pensamiento adscrito como una relación entre el estado de cosas objetivo (o extensión) representado por el contenido del pensamiento y el hablante que tiene dicho pensamiento.⁹ Si, por ejemplo, Ana cree que Juan es soltero, esta creencia interpretada *de re* podría expresarse del siguiente modo: de

⁷ Debemos hacer algunas aclaraciones terminológicas. Mediante el término “referente” hablamos de las entidades, por regla general, extra-lingüísticas que designan nuestras expresiones lingüísticas, en especial los términos singulares. Aunque la referencia alude, más bien, a la relación que existe entre las expresiones y las entidades designadas por ellas, muy a menudo emplearemos este término como sinónimo de “referente”. En lo que sigue, preferimos reservar el uso del término “extensión” para hablar de las entidades extra-lingüísticas (conjuntos o individuos) que designan los términos generales, aunque ello no impide que denominemos a los términos singulares que se refieren al mismo individuo “términos extensionalmente equivalentes” en el sentido de Carnap.

⁸ Hay que decir, no obstante, que Carnap no clarificó tanto como sería deseable de qué manera han de entenderse los conceptos individuales. Ante ello, nosotros sólo recogemos la breve caracterización que ofrece este autor de los conceptos individuales en la obra mencionada, es decir, como los conceptos que constituyen la intensión de un término singular.

⁹ Aunque Carnap considera que la intensión y la extensión son propiedades semánticas de las expresiones lingüísticas, nosotros trasladaremos esta dicotomía al terreno de los pensamientos. Este uso es legítimo debido a que, por una parte, dichos estados mentales poseen un contenido de carácter proposicional, siendo las proposiciones entidades intensionales, y, por otra, remiten a objetos y estados de cosas reales o posibles, que consideramos la extensión de nuestros contenidos.

Juan, Ana cree que es soltero. Con una terminología semi-formal, podríamos describir el pensamiento de Ana de este modo:

(5) $\exists x (x \text{ es Juan} \wedge \text{Ana cree que } x \text{ es soltero})$

Este informe afirma que existe un 'x' tal que 'x' es Juan y respecto del cual Ana cree que es soltero. La expresión latina *de re* significa “de la cosa” o “sobre la cosa” y hace alusión, precisamente, al hecho de que el pensamiento de Ana interpretado *de re* consiste en una actitud mental dirigida hacia la cosa misma (el 'x' que es Juan), acerca de la cual Ana piensa de un determinado modo (cree que él es soltero).

Por su parte, la interpretación *de dicto* de las atribuciones considera a los pensamientos en tanto que están dirigidos hacia un contenido que, debido a su carácter proposicional, tiene una naturaleza intensional. Así, el ejemplo anterior puede ser interpretado *de dicto* adscribiendo a Ana una creencia sobre una proposición, a saber, la proposición de que Juan es soltero. La descripción semi-formal de este pensamiento quedaría entonces del siguiente modo:

(6) Ana cree que $\exists x (x \text{ es Juan} \wedge x \text{ es soltero})$

Puede apreciarse que en la interpretación *de re* ejemplificada en (5) el verbo “creer” está dentro del alcance del cuantificador existencial, dado que dicho cuantificador se aplica a las proposiciones abiertas “x es Juan” y “Ana cree que x es soltero”, en la segunda de las cuales figura dicho verbo.¹⁰ Asimismo, en (5) la variable 'x', que figura también dentro del alcance del verbo “creer”, está ligada por el cuantificador existencial, que permanece fuera de dicho alcance. En cambio, en la interpretación *de dicto*, descrita en (6), tanto el cuantificador existencial como la variable 'x' ligada por él están dentro del alcance del verbo “creer”.

Para aclarar qué consecuencias se desprenden de tomar un pensamiento bajo una u otra interpretación, es conveniente examinar una peculiaridad de las interpretaciones *de dicto* de los pensamientos, a saber, el fallo de la sustituibilidad *salva veritate* de conceptos involucrados en el pensamiento por otros conceptos extensionalmente equivalentes.¹¹ Este fallo supone que cuando llevamos a cabo una sustitución de un

¹⁰ Denominamos “proposición abierta” a lo expresado por una oración abierta en sentido quineano, es decir, por una oración en la que figura alguna variable de individuo no ligada por un cuantificador (Quine 1960, p. 137).

¹¹ Debe señalarse que esta peculiaridad, al igual que el par intensión-extensión, tiene un origen lingüístico que pretendemos trasladar al ámbito psicológico. De hecho, el fallo de la sustituibilidad es propio de los contextos lingüísticos intensionales, de los que constituyen una subclase los contextos de atribución de pensamientos. El uso de operadores modales como “necesario” y “contingente” también genera contextos intensionales.

término que figura en el informe de un pensamiento interpretado *de dicto* por otro término que posee el mismo referente o extensión, puede variar el valor de verdad de dicho informe. Pongamos un ejemplo. En una conversación sobre cine, David afirma que Greta Garbo fue nominada a los premios Óscar por su interpretación en la película *Ninotchka*. Parece natural ofrecer un informe de las creencias de David como el siguiente:

(7) David cree que Greta Garbo fue nominada al Óscar por su participación en la película *Ninotchka*.

La verdad de este informe es patente a la luz de las declaraciones del propio sujeto. Ahora podemos ofrecer una interpretación *de dicto* de la creencia de David, cuya reescritura semi-formal podría ser la siguiente:

(8) David cree que $\exists x$ (x es Greta Garbo \wedge x fue nominada al Óscar por su participación en la película *Ninotchka*).

Supongamos que David viaja a la capital de Suecia y en uno de sus paseos por la ciudad repara en la presencia de una pequeña placa conmemorativa del nacimiento de Greta Lovisa Gustafsson, es decir, de la mujer que adoptaría el nombre de “Greta Garbo” en 1924. David piensa, erróneamente, que esta actriz es natural de los EE.UU. y, por tanto, cree que Greta Lovisa Gustafsson y Greta Garbo son dos personas distintas. Si llevamos a cabo la sustitución del nombre propio “Greta Garbo” en el informe (7) por el nombre propio de la misma actriz, “Greta Lovisa Gustafsson”, obtendremos un nuevo informe:

(9) David cree que Greta Lovisa Gustafsson fue nominada al Óscar por su participación en la película *Ninotchka*.

Sin embargo, por hipótesis, David ignora que ambos nombres propios, “Greta Garbo” y “Greta Lovisa Gustafsson”, se refieren a una y la misma persona. Por ello, no podemos establecer la verdad del informe (9) a partir de la verdad del informe (7), aun cuando el único cambio operado sobre el informe (7), que es un informe verdadero, es la sustitución de un término por otro extensionalmente equivalente. No ocurriría lo mismo si nuestra interpretación de la atribución de la creencia de David expuesta en (7) fuese una interpretación *de re*. En este caso, podríamos reescribirla como:

(10) $\exists x$ (x es Greta Garbo \wedge David cree que x fue nominada al Óscar por su participación en la película *Ninotchka*).

Este informe afirma que, del individuo que es Greta Garbo, David cree que fue nominada al Óscar por su participación en la película *Ninotchka*. (10) es, como (7) y (8),

un informe verdadero. Si decidiéramos llevar a cabo en este caso una sustitución del término “Greta Garbo” por otro término correferencial, “Greta Lovisa Gustafsson”, no obtendríamos las mismas consecuencias que en el caso anterior (un cambio en el valor de verdad del informe). La reescritura semi-formal de tal situación sería la siguiente:

(11) $\exists x (x \text{ es Greta Lovisa Gustafsson} \wedge \text{David cree que } x \text{ fue nominada al Óscar por su participación en la película } \textit{Ninotchka})$.

Tengamos en cuenta que la asignación a la variable ‘x’ del individuo Greta Garbo en esta interpretación *de re* del pensamiento de David figura fuera del alcance del verbo “creer”. Por esto, para que el informe (11) sea verdadero sólo se requiere que exista un ‘x’ tal que ‘x’ es Greta Lovisa Gustafsson y del cual David cree que fue nominada al Óscar por su participación en la película *Ninotchka*. Que nuestro protagonista crea que el ‘x’ nominado al Óscar es Greta Lovisa Gustafsson no forma parte de las condiciones de verdad del informe. En consecuencia, el informe (11) es verdadero.¹²

A la vista de este efecto que se desprende de los distintos modos de interpretar un pensamiento, podemos avanzar que las interpretaciones *de re* de nuestros pensamientos no son apropiadas para asegurar la racionalidad del sujeto.¹³ La importancia de esta conclusión es crucial teniendo en cuenta que a lo largo de nuestra investigación se tomará como norma interpretar a los sujetos ajustándonos en lo posible al denominado “Principio de Caridad”.¹⁴ Dicho principio nos insta a considerar a los hablantes como máximamente racionales y capaces de adecuarse *a priori* a las leyes de la lógica. El Principio de Caridad constituirá, por ello, una piedra de toque para una buena teoría sobre nuestros pensamientos y sus contenidos, de forma que si una explicación acerca de los pensamientos de los sujetos nos acerca a la conclusión de que dichos sujetos mantienen de manera generalizada pensamientos inconsistentes, entonces se tratará de una teoría inadecuada.¹⁵ Ahora bien, como decimos, las interpretaciones *de re* de los

¹² Como hemos indicado en la nota anterior, en los contextos modales se produce un fenómeno análogo. Para verlo, tomemos en consideración el enunciado verdadero: “necesariamente, cuatro es el cuadrado de dos”. Si llevamos a cabo una sustitución del término “dos” por otro término con la misma referencia en el mundo actual, por ejemplo, “el número favorito de Laura”, obtendremos el siguiente enunciado “necesariamente, cuatro es el cuadrado del número favorito de Laura”. Su falsedad es patente, ya que el número favorito de Laura podría haber sido un número distinto de 2.

¹³ Cf. Boghossian (1994), p. 39.

¹⁴ A pesar de que la idea de respetar este Principio en una situación de traducción radical es una idea de Quine (1960), el Principio de Caridad fue denominado como tal y cobró mayor desarrollo en algunas de las contribuciones filosóficas de Davidson sobre la interpretación radical. Vid., por ejemplo, Davidson (1967), p. 27; (1974), p. 197 y (1983), p. 148-149. La comprensión del Principio de Caridad que estamos manejando es un tanto más fuerte que la del propio Davidson. Cf. Davidson (2001), p. XVII.

¹⁵ Quine lo expresa del siguiente modo: “la estupidez de nuestro interlocutor, más allá de cierto punto, es menos probable que una mala traducción” (1960, p. 59). Llevando esta afirmación a nuestro terreno, si

pensamientos de los agentes a menudo arrojan resultados de esa clase. Para comprobarlo, volvamos al ejemplo de David en (6). Podíamos inferir de sus declaraciones que cree que Greta Garbo recibió una nominación a los Óscar por su interpretación en *Ninotchka*, pero como ignora que el nombre real de la actriz sueca era “Greta Lovisa Gustafsson”, si le hiciéramos una pregunta del tipo “¿crees que Greta Lovisa Gustafsson fue nominada a los Óscar por su participación en *Ninotchka*?” es de suponer que David respondería negativamente o expresaría sus dudas al respecto. Ahora bien, al interpretar *de re* sus pensamientos, obtenemos que hay un cierto individuo del cual David cree y no cree (o duda) que recibió una nominación por su interpretación en *Ninotchka*, lo cual quebranta el Principio de Caridad. Nuestro sujeto mantendría pensamientos que no se adecuan *a priori* a las leyes de la lógica, ya que sus pensamientos (interpretados *de re*) sólo serían consistentes si David adquiriera cierta información empírica, a saber, que el nombre real de la actriz sueca conocida como Greta Garbo era “Greta Lovisa Gustafsson”.

No obstante, en caso de que los informes sobre sus creencias recibieran una interpretación *de dicto*, David no sería considerado irracional, dado que no mantendría actitudes incompatibles acerca de *lo mismo*. Para entender esto, recordemos que una interpretación *de dicto* de un pensamiento supone considerar que el agente mantiene una cierta actitud psicológica, del signo que sea, hacia un contenido proposicional. Por ello, si consideramos que los contenidos de los pensamientos que atribuimos a David (cuando sostiene el hecho relevante sobre Greta Garbo y lo niega respecto de Greta Lovisa Gustafsson) son distintos, no podremos decir, entonces, que este agente alberga actitudes incompatibles hacia lo mismo (como sí ocurría en la interpretación *de re* de sus pensamientos). Dicho esto, de ahora en adelante y mientras no indiquemos lo contrario, siempre que hablemos de informes sobre los pensamientos de los sujetos consideraremos que estas actitudes han de ser interpretadas *de dicto*, esto es, como actitudes dirigidas hacia un contenido proposicional que viene especificado en nuestros informes mediante las cláusulas-que.

Antes de exponer algunos de los interrogantes filosóficos que suscitan los pensamientos, es preciso tomar en consideración una serie de características de sus contenidos, así como esclarecer la naturaleza de la relación que existe entre estos contenidos, los sujetos que los piensan y el mundo que representan. Para ello,

una teoría del contenido nos lleva a tomar a los individuos como estúpidos, la estupidez de los individuos es menos probable que la inadecuación de la teoría.

acudiremos a las propuestas de Moya (2004, pp. 151 y ss.), quien enumera y explica tres marcas o características del contenido psicológico. En primer lugar, es plausible considerar que *los contenidos de los pensamientos de los agentes tienen un papel explicativo en su conducta*. Volviendo a nuestro primer ejemplo, el hecho de que María piense que un triunfo del partido P.O. en la oposición sería nefasto para el futuro de la nación explica que este hablante no acuda a los mítines que ofrecen los candidatos de dicho partido, que no haga uso de medios de información afines al partido P.O., que su voto sea favorable al partido P.G. en el gobierno, etc. Parece existir, por tanto, un fuerte nexo entre lo que uno piensa y cómo se comporta. En segundo lugar, *el carácter representacional de los contenidos mentales dota a éstos de propiedades semánticas*. Esta segunda marca del contenido psicológico toma pie en el vínculo existente entre el contenido y aquellos aspectos objetivos (generalmente del mundo externo) que el contenido representa. Así, hemos dicho que el contenido de los pensamientos tenía un carácter proposicional y que, en consecuencia, es susceptible de recibir un valor semántico. Es decir, que si las cosas son tal y como son representadas en el contenido, dicho pensamiento será verdadero. Si, por el contrario, el estado de cosas representado en el contenido no se da, entonces hablaremos de un pensamiento falso.¹⁶ En tercer y último lugar, usualmente se asume que *el sujeto conoce de manera privilegiada los contenidos de sus pensamientos*. Cuál sea el alcance y la caracterización precisa de ese privilegio es algo que tendremos ocasión de discutir a lo largo de este trabajo. Sin embargo, parece intuitivo admitir que el conocimiento que uno tiene del contenido de sus pensamientos es más fiable y seguro que el conocimiento que puede tener del contenido de los pensamientos ajenos. De esta manera, mientras que para atribuir pensamientos a los demás necesitamos realizar inferencias a partir de su comportamiento y sus declaraciones verbales, no parece que por regla general necesitemos esta clase de procedimientos para adscribirnos pensamientos a nosotros mismos. Las atribuciones a terceros toman pie en evidencias empíricas y en la presunción, no siempre correcta, de que los demás son sinceros en sus actos y declaraciones verbales. No obstante, el acceso a los contenidos de nuestros propios pensamientos usualmente está libre de esta clase de obstáculos y, salvo casos de

¹⁶ No hay que confundir el valor de verdad de los pensamientos con el valor de verdad de los informes que realizamos sobre los pensamientos de los sujetos. Por ejemplo, podemos pensar en Juan, que sufre de paranoia, de quien es verdadero el informe “Juan cree que todos sus vecinos le observan”. No obstante, puesto que Juan en realidad no despierta en absoluto la curiosidad de sus vecinos, la creencia de Juan será falsa al serlo su contenido.

autoengaño, presuponemos una fiabilidad especial en los datos que dicho acceso nos proporciona. Una buena teoría del contenido psicológico debería respetar este hecho de la *accesibilidad privilegiada*.¹⁷

Los pensamientos y sus contenidos son, como decimos, objeto de un considerable interés filosófico. Por ejemplo, es una cuestión sumamente interesante si los pensamientos, en tanto que entidades de carácter mental, tienen cabida en la ontología presupuesta en las ciencias físico-naturales. Por supuesto, las respuestas a esta pregunta tienen que ver con el estatus ontológico que otorguemos a lo mental como distinto, reducible o idéntico a lo físico. Otra cuestión, en cierta medida relacionada con la anterior, tiene que ver con el grado de receptividad que una psicología científica debe presentar ante los informes ordinarios de la psicología popular. Estos problemas, cuya importancia debe destacarse, no nos conciernen en esta investigación. El aspecto que trataremos principalmente en nuestro trabajo es la pregunta por las condiciones de identidad de los pensamientos de los hablantes y, muy especialmente, *la pregunta por los factores de los cuales dependen constitutivamente sus contenidos*. En este sentido, hemos dicho que la identidad de un pensamiento depende del tipo de actitud que adopta el sujeto hacia el contenido del pensamiento y de la identidad del propio contenido. No obstante, en lo sucesivo el constituyente de nuestros pensamientos que nos preocupará de manera principal será su contenido y, en consecuencia, dedicaremos toda nuestra atención a la pregunta por las condiciones de identidad de nuestros contenidos mentales, sin detenernos en la pregunta por las condiciones de identidad de los distintos tipos de actitudes psicológicas. De ello se siguen dos consecuencias. En primer lugar y con el fin de simplificar la discusión, la mayoría de los ejemplos que figuran en el resto de nuestro trabajo constituyen informes de creencias, aunque éstas sean sólo un subconjunto de los pensamientos. En segundo lugar y en aras de la brevedad, caracterizaremos ciertas teorías psicológicas (el internismo y el externismo psicológicos) en tanto que teorías que se ocupan de la individuación de nuestros pensamientos, aunque ambas versen, más precisamente, sobre la individuación *del contenido* de nuestros pensamientos.

Para arrojar la mayor claridad posible sobre el problema que nos ocupa y evitar con ello algunos malentendidos, debemos distinguir entre la dependencia *causal* y la

¹⁷ Es necesario aclarar que cuando hablamos de accesibilidad privilegiada nos estamos refiriendo a la clase de conocimiento que uno tiene sobre sus propios contenidos mentales y sobre sus pensamientos en tanto que éstos pueden figurar como contenido de otros pensamientos. En este sentido, sé con certeza y de manera privilegiada qué es lo que pienso, independientemente de que el contenido de lo que creo sea verdadero, y, asimismo, sé con certeza y de manera privilegiada que pienso lo que pienso, independientemente de que lo que pienso sea verdadero.

dependencia *constitutiva* de los pensamientos.¹⁸ Cuando hablamos de que los pensamientos dependen causalmente de ciertos factores, estamos aludiendo al hecho de que dichos estados mentales se han originado a partir de elementos con los que el sujeto ha establecido una vinculación causal más o menos directa. El pensamiento de Antonio acerca de la frialdad del agua de su bañera depende causalmente de sus percepciones sensoriales de la temperatura del agua. El hecho de que Eva piense que la invasión musulmana de la Península Ibérica tuvo lugar en el 711 d.C. depende causalmente de la información transmitida por sus profesores y contenida en los libros de historia que ha leído. No obstante, cuando hablamos de la dependencia de la constitución de un pensamiento respecto de ciertos elementos estamos tomando partido en una *cuestión de carácter metafísico*. Decimos que una entidad A depende constitutivamente de una segunda entidad B si A no sería A de no ser por B. Esto es, el carácter de la entidad A *en tanto que* A depende de su vinculación con B. Con frecuencia emplearemos los términos “identidad” o “carácter” de un pensamiento A de manera intercambiable, teniendo en cuenta que mediante el uso de estos términos nos estamos refiriendo a lo mismo, a saber, aquello que hace que A sea el pensamiento que es y no cualquier otro. Si ese pensamiento A depende constitutivamente de B entonces será este factor, B, el que hará que A sea el pensamiento que es, es decir, que tenga la identidad o el carácter que tiene.

Una noción que suele aparecer a menudo en los debates sobre las condiciones de identidad de nuestros pensamientos y sus contenidos es la noción de *sobreveniencia*.¹⁹ Vamos a explicar su significado con el fin de aportar el mayor sentido al debate que nos ocupa. Tomemos en consideración dos conjuntos de propiedades, el conjunto de propiedades A y el conjunto de propiedades B. Decimos que el conjunto de propiedades A sobreviene al conjunto de propiedades B (también denominado “base de sobreveniencia”) si y sólo si dos objetos o individuos no pueden diferir en sus propiedades A sin diferir en sus propiedades B. Es decir, si dos individuos u objetos tienen exactamente las mismas propiedades B, entonces es necesario que tengan las mismas propiedades A. Pongamos un ejemplo para aclarar esta relación. Según reza la definición de velocidad de un cuerpo, ésta depende de la proporción entre el espacio que dicho cuerpo recorre y el lapso de tiempo que tarda en hacerlo. Un cuerpo que recorre 6 metros en 3 segundos tiene exactamente la misma velocidad que un cuerpo que recorre

¹⁸ Cf. Wilson (2003), pp. 256-257; Davies (2000), p. 322 y Moya (2004), pp. 153-154.

¹⁹ Cf. Brown (2004), p. 4.

4 metros en 2 segundos, a saber, la velocidad de 2 metros por segundo. La relación de sobrevenida se aplicaría en este caso tomando, por un lado, la velocidad como propiedad A y, por otro, la proporción entre el espacio recorrido y el tiempo empleado como propiedad B. Si dos cuerpos difieren en sus propiedades A (sus respectivas velocidades) ello implica, necesariamente, una diferencia en sus propiedades B (las respectivas proporciones entre el espacio recorrido y el tiempo empleado). Para aplicar la relación de sobrevenida al ámbito que nos ocupa, que son los pensamientos y su dependencia constitutiva respecto de otros factores, consideraremos que los estados mentales de carácter proposicional que atribuimos a los sujetos pueden tomarse en tanto que propiedades de dichos sujetos. La cuestión acerca de la dependencia constitutiva de los pensamientos planteada en términos de sobrevenida concibe la posesión del pensamiento de que *P* (y no cualquier otro) como la propiedad A y pretende determinar cuál es la propiedad o, más bien, el conjunto de propiedades B a las cuales sobreviene la propiedad A. Dicho de otro modo, si consideramos que Juan y Pedro ameritan la atribución de creencias distintas (Juan posee la propiedad A y Pedro no), entonces habrá algo distinto en ambos escenarios de atribución²⁰ (una diferencia en las propiedades B) en virtud de lo cual establecemos una diferencia en sus respectivos estados psicológicos. La pregunta por las condiciones de identidad de los pensamientos pretende descubrir cuáles son los rasgos de los escenarios de atribución que constituyen la base de sobrevenida de estos estados mentales.²¹

Para finalizar con esta introducción sería conveniente relacionar las observaciones que acabamos de presentar con el título de nuestro trabajo. De este modo, el problema que nos ocupará en lo sucesivo es la discusión sobre *la posibilidad de compatibilizar el externismo psicológico con el acceso privilegiado a nuestros propios pensamientos*. Aunque reservaremos el próximo capítulo para clarificar al máximo el contenido de las tesis externistas, por el momento podemos indicar que el externismo psicológico es una posición filosófica acerca de las condiciones de identidad de los contenidos de nuestros estados mentales de carácter proposicional. Concretamente, los partidarios de esta posición defienden que algunas de dichas condiciones son externas a nuestra mente y,

²⁰ Denominamos “escenario de atribución” al conjunto que incluye tanto las propiedades intrínsecas de los individuos (su configuración neuronal y su fenomenología) como las propiedades externas a dichos sujetos (los géneros naturales presentes en su entorno y los usos lingüísticos de su comunidad).

²¹ A partir de ahora consideraremos que preguntar por las *condiciones de identidad* de un pensamiento equivale a preguntar por las *condiciones, elementos o factores que intervienen en su individuación*. De este modo, si pretendemos averiguar qué hace que el pensamiento de que P tenga esa identidad (que sea el pensamiento de que P y no de que Q o R) lo que se buscará en esta investigación son los factores que individualizan dicho pensamiento en tanto que pensamiento de que P.

en esta medida, podrían resultar desconocidas para nosotros. La posición rival del externismo psicológico es el internismo, que niega el carácter externo de cualquiera de los elementos que se requieren para determinar la identidad de nuestros contenidos mentales.

Tras esta breve caracterización de la oposición internismo/externismo psicológico, diremos que, en tanto que ambos versan sobre (la identidad de) nuestros pensamientos y contenidos, sería deseable que sus respectivas propuestas respetaran las tres marcas del contenido mental que hemos indicado anteriormente, debido a que se trata de rasgos que gozan de un carácter muy intuitivo. A este respecto, ninguna teoría de la individuación del contenido psicológico (internista o externista) debería violentar la intuición de que dos sujetos que toman en consideración un mismo contenido psicológico con una misma actitud tienden a llevar a cabo conductas muy semejantes, si no idénticas y, viceversa, si dos hablantes presentan una misma actitud psicológica ante contenidos distintos, sus acciones relevantes deberán diferir. Por otra parte, tanto el internismo como el externismo deben poner en juego un modo de determinación de nuestros contenidos que respete su carácter representacional y dé cuenta de las condiciones de verdad que corresponden de manera efectiva a los contenidos mentales así individuados. Por último, una buena teoría del contenido psicológico debería preservar en principio el acceso privilegiado del sujeto al contenido de sus propios pensamientos. Como tendremos ocasión de comprobar, se ha desarrollado un debate muy amplio y complejo sobre la posibilidad de conjugar dicho acceso con la individuación externista de los pensamientos. La importancia de este debate es máxima, ya que si una teoría sobre las condiciones de identidad de los contenidos mentales atenta contra el carácter especial de nuestro conocimiento de los mismos, entonces no puede tratarse de una teoría aceptable.

A lo largo del presente trabajo examinaremos en profundidad los argumentos que se han desarrollado para demostrar tanto la incompatibilidad como la compatibilidad del externismo psicológico con el privilegio del conocimiento que tenemos de nuestros contenidos mentales, siendo la propia aceptabilidad del externismo aquello que, en último término, se dirime en dicha polémica.

Capítulo 1: Externismo semántico y externismo psicológico

El problema acerca de los pensamientos y los contenidos mentales que nos ocupará en este capítulo ha sido introducido en las páginas precedentes. Lo que se busca son aquellos elementos en virtud de los cuales el pensamiento que adscribimos a un sujeto tiene cierto contenido. Según hemos indicado, los autores dedicados a esta cuestión han ofrecido respuestas muy diferentes, aunque es frecuente clasificar sus aportaciones en dos grandes grupos. El primero de ellos, el grupo de los denominados *internistas*, incluye a todos aquellos autores que juzgan que los factores *internos* al hablante, factores con respecto a los cuales el hablante tiene pleno acceso epistémico, constituyen la condición suficiente y necesaria para determinar la identidad de sus contenidos mentales. Sin embargo, otros autores han optado por mantener que, al menos en ocasiones, es preciso contar con algunos elementos *externos* a los sujetos, elementos que tales sujetos podrían desconocer, para determinar qué es lo que creen, desean o esperan. De respaldar esta última opción, nos situaremos en las filas del denominado *externismo psicológico*.

No busquemos, por el momento, un significado más preciso al carácter bien interno, bien externo, de los ingredientes del contenido mental. Antes es recomendable ofrecer una serie de aclaraciones sobre el par de etiquetas internismo-externismo. Varios debates sobre cuestiones relacionadas con el lenguaje, la mente y el conocimiento se encuentran polarizados en opciones internistas y externistas, lo cual indica que no tiene demasiada utilidad declararse internista o externista a secas, sino que uno suscribe alguno de estos dos enfoques con respecto a tal o cual cuestión. Nosotros hablaremos de *internismo* y *externismo psicológicos* para encauzar la polémica que concierne al *carácter interno o (parcialmente) externo de las condiciones de identidad de los contenidos de nuestros pensamientos*. Con esta decisión nos separamos de un hábito con cierto arraigo, a saber, el empleo del rótulo “internismo-externismo semántico” para tratar el problema de la determinación de nuestros contenidos psicológicos. Como veremos a lo largo de este capítulo, este hábito se apoya en un hecho bien establecido: las teorías externistas del contenido psicológico han derivado históricamente de las teorías externistas en semántica y han recibido apoyo teórico de sus conclusiones. Con todo, es conveniente separar ambas cuestiones desde un principio y, por ello, cuando se hable del *internismo o externismo en semántica* lo haremos para hacer referencia a la

disputa en torno a la determinación bien interna, bien en parte externa, de las propiedades semánticas de nuestras expresiones lingüísticas.

Iniciamos, pues, este capítulo con el objetivo principal de examinar las propuestas externistas en semántica, cuya denominación genérica será *Nuevas Teorías de la Referencia*, ya que tales propuestas constituyen la base teórica del externismo psicológico.²² Llevaremos a cabo dicho examen en cuatro secciones. A lo largo de la primera sección abordaremos un primer punto de contraste entre el internismo y el externismo semántico, tomando en consideración las tesis internistas clásicas de Frege y Russell en torno al *valor semántico* de los nombres propios para contraponerlas a las ideas externistas que proponen al respecto las Nuevas Teorías de la Referencia. En la segunda sección nos centraremos en otro punto de contraste, a saber, aquel que concierne a *la determinación de la referencia* de determinadas expresiones, especialmente los nombres propios y los términos de género natural. Más en particular, en esa sección comenzaremos exponiendo las críticas de Kripke a la teoría clásica de Frege y Russell acerca de los nombres propios y examinaremos qué imagen alternativa, de carácter externista, nos propone este autor para dar cuenta del vínculo referencial, tanto de esa clase de expresiones como de los términos de género natural. Igualmente, en la segunda sección también se abordarán las ideas de Putnam acerca de las propiedades semánticas de los términos de género natural, dedicando especial atención a los experimentos mentales que este autor plantea con el objetivo de mostrar las insuficiencias del enfoque internista acerca de la determinación de la referencia de los términos de género natural. En la tercera sección propondremos una reelaboración de los experimentos mentales mencionados, de modo que sus conclusiones sirvan para reforzar las intuiciones del externismo psicológico. En la cuarta y última sección analizaremos los experimentos mentales de un segundo autor, Tyler Burge, elaborados *ex profeso* para ofrecer sustento teórico al modo externista de individuación de nuestros contenidos mentales.

²² Cabe pensar que resulta un tanto inadecuado calificar de “Nuevas” a una serie de propuestas semánticas que comenzaron a ser enunciadas hace ya cuarenta años. Sin embargo, esta denominación no es del todo inusual (cf. Perry 1988; Smith 1995; Wettstein 2004 y Martí 1995, 2003) y presenta dos ventajas. La primera es que su novedad se establece en contraste con el carácter clásico del descriptivismo de Frege y Russell y la segunda es que presenta cierta neutralidad en relación a las dos familias de teorías que situamos en su seno, a saber, la Teoría de la Referencia Directa y la Teoría Histórico-Causal de la Referencia.

1. *Externismo semántico (I): La Teoría Clásica y la Teoría de la Referencia Directa*

La clase de externismo psicológico que constituye el centro de nuestras preocupaciones en este trabajo concierne a aquellos pensamientos que expresamos mediante preferencias en las que figuran términos generales, esto es, términos que designan conjuntos de entidades (“agua”, “oro”, “artritis”, “carburador”).²³ Sin embargo, para proporcionar pleno sentido a las tesis del externismo psicológico es conveniente tomar en cuenta, en primer lugar, la polémica existente entre los partidarios del internismo y del externismo en el ámbito de la semántica de los términos singulares y, especialmente, de los nombres propios. Las razones que nos asisten para proceder de esta manera son de índole histórica y metodológica. En primer lugar, el debate relativo al comportamiento semántico (externista o internista) de los términos singulares se ha suscitado con anterioridad a la disputa sobre las condiciones de identidad (externas o internas) de los contenidos psicológicos expresados mediante preferencias en las que figuran términos generales. En segundo lugar, el análisis de la primera discusión nos permitirá presentar ciertas nociones técnicas y desplegar un conjunto de problemas que serán de interés a lo largo de todo el trabajo. Por estos motivos, dedicamos la presente sección al análisis de la discusión entre los defensores de la semántica clásica de corte internista (Frege y Russell, principalmente) y los partidarios externistas de las Nuevas Teorías de la Referencia,²⁴ centrándonos en el problema de la contribución proposicional de los nombres propios.

1.1 *Los términos singulares y sus propiedades semánticas*

La mejor manera de adentrarnos en este debate pasa por explicar, primeramente, qué entendemos por términos singulares, un subconjunto de los cuales son los nombres propios, y, en segundo lugar, cuáles son sus propiedades semánticas. Pues bien, por lo que respecta a lo primero, diremos que los *términos singulares* son expresiones

²³ En la introducción a nuestro trabajo señalamos la coexistencia de dos concepciones diferentes acerca del referente o extensión de los términos generales (o bien como un conjunto de individuos o bien todos y cada uno de los individuos que pertenecen a ese conjunto). Según se dijo, tomamos la decisión de atender únicamente a la primera de estas concepciones.

²⁴ No deja de ser una ironía que las *Nuevas Teorías de la Referencia* tengan, en el fondo, un carácter “reaccionario” y muy poco novedoso, como dice Salmon en (2005a, p. 9), al señalar el hecho de que sus desarrollos toman pie en una teoría más tradicional acerca de los nombres propios que la teoría fregeana, a saber, la teoría de J. S. Mill.

lingüísticas cuya función es introducir en el discurso a un único individuo. Los nombres propios (“Juan”, “Esteban”), los términos indéxicos singulares y las descripciones definidas singulares son los tipos de expresiones que suelen clasificarse como términos singulares. La peculiaridad de los *términos indéxicos* consiste en que su referente, una entidad singular, varía en función del contexto de preferencia, contexto que incluye, al menos, el hablante, el oyente, así como el tiempo y el lugar de emisión. Algunos ejemplos de términos indéxicos son “hoy”, “te”, “yo”, “esto”, “ahora”. Como puede verse, en la oración “te quiero” el referente del pronombre “te” varía en función de si es Eva quien se dirige a Ricardo mediante su uso o de si es Ricardo quien profiere esta oración para referirse a Eva. Por lo que respecta a las *descripciones definidas singulares*, se trata de términos singulares²⁵ que comienzan por el artículo determinado en singular, que expresan una propiedad que pretende ser poseída por un único individuo y cuya referencia viene determinada por la referencia de sus expresiones componentes. “El autor de *Crimen y Castigo*”, “el hijo de Sofronisco” y “la jugadora más alta de la selección española de baloncesto en el *Eurobasket* de 2009” son descripciones definidas singulares que se refieren, respectivamente, a Fiódor Dostoievski, Sócrates y Tamara Abalde.

En cuanto a las propiedades semánticas de estos términos que ocupan a los defensores del internismo y del externismo, diremos que tales propiedades son la *referencia* y el *significado*, entendido este último como valor semántico.²⁶ En lo que respecta a la primera propiedad, estas teorías discuten en virtud de qué un término singular designa al individuo que constituye su referente o, lo que es igual, se busca delimitar la naturaleza del vínculo referencial entre la expresión y la entidad a la que se refiere. Mientras que los partidarios del internismo sitúan *todos* los factores que determinan la referencia de un término singular en el ámbito cognoscitivo (“interno” al sujeto), los externistas defienden que es necesario tomar en consideración, *además*, elementos que residen en el entorno natural o social de los usuarios del lenguaje (“externos” a cada uno de estos usuarios). Siguiendo la convención propuesta por Martí

²⁵ Como veremos, Russell no considera a las descripciones definidas como términos singulares.

²⁶ Hemos elegido separar la problemática de la determinación de la referencia de la relativa al valor semántico de los términos aun cuando ambas son fundamentales para explicar en qué consiste el externismo semántico que pusieron en boga las Nuevas Teorías de la Referencia. En palabras de Martí, “es impropio hablar de la Revolución en la Filosofía del Lenguaje [en tanto que una única revolución iniciada por las Nuevas Teorías de la Referencia], dado que hay, al menos, dos diferentes revoluciones. Ambas son revoluciones levantadas en contra de la aproximación fregeana y en contra de la semántica tradicional intensional, pero son revoluciones diferentes” (Martí 2003, p. 174).

en (1995), emplearemos la denominación de *Millianismo* para designar aquel grupo de propuestas que, dentro de las Nuevas Teorías de la Referencia, rechazan que el vínculo referencial entre nuestras palabras y los aspectos del mundo que designan venga determinado por las descripciones que los usuarios del lenguaje asocian con aquéllas y que, por tanto, la conexión entre los términos y los objetos tenga un carácter mediato en este sentido. Sin embargo, dejaremos los detalles de esta disputa para la segunda sección del capítulo.

Lo que constituye nuestra preocupación en la sección presente es, más bien, la discusión entre internistas y externistas acerca de la naturaleza de la segunda propiedad mencionada: el valor semántico de los términos singulares y, especialmente, de los nombres propios. Para explicar qué se entiende por valor semántico, resulta conveniente apelar a la noción de proposición o contenido proposicional expresado por nuestras oraciones. Ya en la introducción a nuestro trabajo denominamos “proposición” a lo que nuestras oraciones dicen, esto es, a (i) la entidad abstracta expresada por dichas oraciones, que (ii) posee valor de verdad y que (iii) lo posee primariamente. Cuando decimos que el internismo y el externismo semántico se ocupan de la contribución semántica de ciertas expresiones, nos estamos refiriendo al modo en que éstas contribuyen a determinar cuál es la proposición expresada por esa oración. La aportación peculiar de cada término a este respecto recibe el nombre de *valor semántico*. Por ejemplo, “Juan” y “Esteban”, dos nombres propios que usamos para referirnos a individuos distintos, son términos que aportan un valor semántico diferente a las proposiciones expresadas por las oraciones en las que aparecen. Para entenderlo mejor, podemos emplear estas dos oraciones:

(1) Esteban ocupa un banco.

(2) Juan ocupa un banco.

Cabe apreciar que las oraciones (1) y (2) no dicen *lo mismo*, es decir, que la proposición que expresamos al proferir cada una de estas oraciones es una proposición diferente. En el primer caso, la oración “Esteban ocupa un banco” sería verdadera sólo si Esteban, de hecho, ocupa un banco, mientras que “Juan ocupa un banco” expresaría una proposición verdadera sólo si es Juan quien lo ocupa. Mediante esto se comprueba que la diferencia en las proposiciones que expresan (1) y (2) se establece a partir del distinto valor semántico de los nombres propios que figuran en ellas.²⁷

²⁷ Dicho todo esto, hay que realizar una observación. La diferencia entre las aportaciones semánticas de estas expresiones es una diferencia en los *tipos* de aportaciones que realizan los distintos *tipos* de

De nuevo, internistas y externistas ofrecen explicaciones incompatibles en torno al valor semántico de los términos singulares. A este respecto, los defensores del internismo tienden a asimilar la contribución proposicional de un término singular a entidades de carácter “interno” al sujeto, típicamente, la *información descriptiva* que el usuario de un término singular asocia con éste. Los partidarios del externismo consideran, por su parte, que la contribución de la mayoría de los términos singulares (los nombres propios, términos indécicos y determinados usos de las descripciones definidas) a las proposiciones expresadas por las oraciones en las que figuran es idéntica al *referente* de dichos términos. Como la entidad a la que se refieren estas expresiones es, por lo general, una entidad del mundo externo, se entiende el carácter “externista” de la contribución proposicional que, según esta teoría, corresponde a tales expresiones. Siguiendo la convención de Martí en (1995), denominaremos *Teoría de la Referencia Directa* al conjunto de propuestas que, incluidas en el seno de las Nuevas Teorías de la Referencia, defienden que la contribución semántica de ciertos términos singulares es *puramente referencial* en el sentido descrito.

1.2 *La Teoría Clásica de Frege sobre el valor semántico de los nombres propios*

Establecida la diferencia entre los problemas de la determinación de la referencia y aquellos que conciernen al valor semántico de los términos, así como la distinción entre las dos familias de propuestas que, dentro de las Nuevas Teorías de la Referencia, se ocupan del estudio de una y otra cuestión, reservaremos el grueso de esta sección primera a examinar los detalles de la polémica sobre el valor semántico de los nombres propios en la que se enfrentan, por un lado, las propuestas clásicas de Frege y Russell y, por otro, la Teoría de la Referencia Directa.

La presente subsección está dedicada en exclusiva a Frege, cuya propuesta fundamental sobre la semántica de los nombres propios fue elaborada por él a partir de un problema presentado en (1892). El problema es el siguiente. Sean “a” y “b” dos nombres propios que designan el mismo individuo; consideremos los siguientes

términos singulares. De no entenderlo así, nuestra búsqueda sería trivial, porque todo el mundo admite que las expresiones “Cicerón” y “el primer hombre que pisó la luna” contribuyen de manera distinta a las proposiciones expresadas por las oraciones “la enciclopedia explica quién fue Cicerón” y “la enciclopedia explica quién fue el primer hombre que pisó la luna”. Lo que se intenta explicar es si dos tipos de términos singulares correferenciales como el nombre propio “Cicerón” y “el autor de las *Catilinas*” ofrecen contribuciones semánticas distintas en tanto que expresiones que pertenecen a clases de términos singulares diferentes (nombres propios y descripciones definidas, respectivamente).

enunciados de identidad verdaderos: “a es a” y “a es b”. Es natural admitir, dice Frege, que estos dos enunciados tienen un distinto valor cognitivo. Por un lado, la verdad de la aseveración “a es a” es conocida *a priori*, esto es, sin necesidad de justificación empírica. Además, según Frege, “a es a” tiene un carácter analítico en el sentido kantiano.²⁸ Sin embargo, no ocurre lo mismo con el enunciado “a es b”. Para verlo, pensemos en dos nombres propios que designan el mismo individuo, por ejemplo, “Fred Astaire” y “Frederick Austerlitz” y formemos a partir de ellos dos enunciados de identidad como los propuestos por Frege:

(3) Fred Astaire es Frederick Austerlitz.

(4) Fred Astaire es Fred Astaire.

Mientras que lo aseverado por el enunciado (4) es una verdad trivial, muchos hablantes pueden negar o dudar de lo afirmado por (3) al desconocer que ambos nombres, “Fred Astaire” y “Frederick Austerlitz”, se refieren al mismo individuo. La diferencia en las actitudes cognitivas que un hablante exhibe ante “a es a” y “a es b” pone de manifiesto que este hablante no comprende lo mismo cuando toma en consideración ambos enunciados. Si esto es así, entonces diremos, con Frege, que “a es a” y “a es b” son oraciones que difieren en cuanto a su *valor cognitivo*, lo cual supone que, si fueran tomadas en consideración por un sujeto que asiente ante la proferencia de la primera oración y niega la verdad de la segunda, éste captaría un *pensamiento* distinto en cada caso. La solución que propone Frege al problema descrito pasa por explicar la diferencia de valor cognitivo entre “a es a” y “a es b” apelando al diferente *sentido* de los nombres propios “a” y “b”. Conviene, por tanto, atender al significado de este término, “sentido”, que tiene un papel central en la propuesta semántica de Frege y examinar, así, qué funciones desempeña esta noción.

Para empezar, diremos que el sentido de una expresión es (i) *un modo de presentación de su referente*: dos expresiones pueden referirse al mismo individuo pero presentándolo de manera distinta. A este respecto, Frege mantiene que el sentido de un nombre propio “ilumina parcialmente” a la entidad designada por él (1892, pp. 25-26), esto es, presenta a su referente como el único individuo que satisface cierta condición.

²⁸ Es discutible que ese enunciado de identidad tenga un carácter analítico según la caracterización de Kant, quien considera que en los enunciados analíticos el concepto del predicado se haya incluido en el concepto del sujeto. Así, en el caso del enunciado “a es a” no cabe distinguir entre sujeto y predicado. No obstante, este enunciado sería perfectamente analítico según la comprensión tradicional de estos enunciados como enunciados verdaderos en virtud exclusivamente del significado de sus términos componentes.

A partir de esto, entenderemos la explicación que la teoría de Frege ofrece a la primera de las preguntas que conciernen al análisis semántico, a saber, que el elemento responsable de explicar el vínculo referencial entre un nombre propio y la entidad que designa es el sentido del nombre, ya que éste nos proporciona un criterio, una condición que debe cumplir el individuo designado por el nombre. Según la manera más usual de entender a Frege, este autor sostendría que el sentido de un nombre propio es idéntico al sentido de una descripción definida, es decir, de una descripción definida que expresaría una propiedad que identifica al individuo al que el nombre se refiere.

Para comprender esta tesis con mayor claridad, acudamos al ejemplo que propone este autor. En (1892, p. 25, n. 2) Frege nos dice que el sentido del nombre propio “Aristóteles” (para un hablante) sería idéntico al de una descripción definida como “el discípulo de Platón y maestro de Alejandro Magno”.²⁹ Aplicando estas pautas a nuestro ejemplo, diríamos que “Fred Astaire” y “Frederick Austerlitz” son nombres que nos presentan a su referente (uno y el mismo individuo) de una manera distinta y, en consecuencia, se trata de expresiones con diferentes sentidos. El sentido de “Fred Astaire” sería idéntico al sentido de una descripción como “el famoso bailarín que fue pareja artística de Ginger Rogers”, mientras que el sentido de “Frederick Austerlitz” vendría dado por una descripción como “el individuo que será conocido popularmente como ‘Fred Astaire’”. Si un hablante desconociera que el nombre “Frederick Austerlitz” conlleva un diferente modo de presentación del (mismo) individuo designado por el nombre propio “Fred Astaire”, esto explicaría sus diferentes actitudes cognitivas ante la preferencia de las oraciones de identidad “Fred Astaire es Fred Astaire” y “Fred Astaire es Frederick Austerlitz”. Por tanto, el distinto sentido de los nombres propios “Fred Astaire” y “Frederick Austerlitz” explica las diferencias en (ii) *el valor cognitivo* de dichos nombres y, más en general, las diferencias en el valor cognitivo de estas oraciones o de cualesquiera oraciones que se diferencien exclusivamente en la sustitución de un nombre propio por otro con distinto sentido pero con la misma referencia. Frege expresa esta idea diciendo que un hablante que tomara en

²⁹ Tanto Frege como Russell admiten que distintos hablantes pueden asociar descripciones definidas diferentes con el mismo nombre propio; en el caso del primero como proporcionando el sentido del nombre propio – vid. nota 30 –, en el caso del segundo – vid. próxima sección – como abreviadas por el nombre propio en cuestión. No obstante, en aras de la brevedad, en nuestra exposición de las teorías de Frege y Russell supondremos por regla general que un nombre propio viene vinculado con una única descripción definida – en principio común para todos los usuarios del nombre propio –.

consideración ambas oraciones de identidad estaría captando pensamientos diferentes en cada caso.³⁰

Es necesario aclarar, no obstante, que el término “pensamiento” tiene un significado un tanto especial para Frege. Los *pensamientos* son los sentidos de las oraciones asertivas, los cuales constituyen entidades objetivas y abstractas que podemos captar mediante nuestra actividad subjetiva de pensar (Frege 1918, p. 54) y que son los portadores primarios de valores de verdad. Si esto es así, podemos apreciar un claro parentesco entre el uso del término “pensamiento” por parte de Frege y nuestro uso del término “proposición”.³¹ Además, el sentido de un enunciado, el “pensamiento” que un sujeto capta al tomar en consideración dicho enunciado, es función del sentido de sus partes componentes y de su estructura. Como vemos, los nombres “Fred Astaire” y “Frederick Austerlitz” tienen distintos sentidos y, por tanto, comprometen la identidad de los pensamientos expresados por las oraciones en las que figuran. Hablando en términos de proposiciones, diremos que, en tanto que “a” y “b” aportan un sentido distinto a los pensamientos expresados en “a es a” y “a es b”, su contribución a las proposiciones que esos enunciados expresan es, asimismo, distinta y determina una diferencia en dichas proposiciones. Hemos denominado a esta clase de aportación “valor semántico” de nuestras expresiones, por lo cual podemos afirmar que el sentido de una expresión equivale a (iii) *su valor semántico*.

Por otro lado, para este autor el sentido de una expresión desempeña una función adicional, a saber, constituye (iv) *la referencia de las expresiones que figuran en la cláusula-que usada para informar del contenido de los pensamientos de un hablante*. Para entender esto, comencemos por señalar que Frege advierte, como hará Russell en (1905),³² que la sustitución de una expresión por otra con la misma referencia en las cláusulas-que de nuestros informes psicológicos conlleva a menudo un cambio en el

³⁰ En (1918), p. 41, Frege dirá que en el caso de que dos hablantes asocien distintas descripciones definidas con un mismo nombre estarán hablando lenguajes distintos con respecto a ese nombre.

³¹ Según vimos, los contenidos de nuestros pensamientos tienen un carácter proposicional, esto es, se trata de entidades abstractas, objetivas y portadoras de valores de verdad. La existencia de similitudes entre esta concepción de los contenidos mentales y los pensamientos fregeanos se debe a que unos y otros poseen las tres propiedades enunciadas.

³² Señalaremos, sin embargo, que para Russell existe cierta clase de nombres propios en relación a los cuales no se presenta este problema. Más adelante veremos qué diferencias existen entre esa clase de nombres propios (los nombres lógicamente propios) y la mayoría de los nombres propios, esto es, los nombres propios en el sentido usual del término. En contextos de oraciones de atribución de actitudes proposicionales únicamente la sustitución de nombres propios correferenciales de esta última clase es la que conlleva, en ocasiones, una variación en el valor de verdad de dicho tipo de oraciones.

valor de verdad del informe en su conjunto. Así, tomemos en consideración un informe sobre los pensamientos de Juan como:

(5) Juan quiere que María se encargue de la alimentación de su gato.

Juan, por su parte, ignora que María odia a los gatos y no pierde la oportunidad de envenenar a todo felino que se cruce con ella. Veamos el siguiente informe:

(6) Juan quiere que la mujer que envenena a todo gato que encuentra alimento a su gato.

La falsedad de (6) es patente, aun cuando la única variación que presenta con respecto al informe original es la sustitución del nombre propio “María” por una expresión que designa el mismo individuo (“la mujer que envenena a todo gato que encuentra”). La explicación que ofrece Frege de este fenómeno se basa en el hecho de que en un contexto indirecto, como los contextos de oraciones de atribución de actitudes proposicionales,³³ las palabras que figuran en la cláusula subordinada (cláusula-que, en nuestro caso) no conservan su sentido y referencia usuales, sino que su referencia es su sentido usual y su sentido, un sentido indirecto, distinto del usual.³⁴ Como resultado de esto, una sustitución de expresiones en un contexto indirecto como la que presentamos no equivaldría a una sustitución de expresiones correferenciales, puesto que en dichos contextos las expresiones no tienen su referencia usual.

A modo de recapitulación, diremos que, para Frege, el sentido de un nombre propio, en tanto que dimensión semántica distinta de su referencia, constituye, por un lado, (i) un modo de presentación del objeto designado por dicho nombre y, en esta medida, un criterio para determinar su referente. El sentido es, además, (ii) el valor cognitivo de un nombre propio, aquello que el nombre aporta a los pensamientos que un agente capta al comprender las oraciones en las que figura el nombre. Por otra parte, el valor cognitivo de un nombre equivale a su (iii) valor semántico, esto es, a su contribución a la proposición expresada por una oración que contiene dicho término, en la medida en que consideramos que esa proposición puede ser identificada con el pensamiento que el hablante capta al entender la oración. Por último, el sentido de un nombre propio es idéntico a (iv) la referencia del nombre en su aparición en una

³³ Nótese que en la denominación usual de estos contextos en tanto que *contextos de atribución de actitudes proposicionales* la expresión “actitud proposicional” es equivalente a “pensamiento”. Hecha esta precisión y respetando la decisión terminológica recogida en la nota 2 de la introducción, preferimos reservar el término “actitud” para referirnos al modo en que un sujeto toma en consideración un contenido psicológico y, por tanto, para referirnos a un componente de los pensamientos.

³⁴ En (1892), p. 33, Frege habla de cláusulas subordinadas introducidas por los verbos “dice”, “oye”, “es de la opinión de”, “está convencido de”, “concluye” y otros similares.

cláusula-que de una expresión que informa sobre los estados mentales de carácter proposicional de un sujeto.

1.3 *La Teoría Clásica de Russell: descripciones, nombres lógicamente propios y distinción entre conocimiento por familiaridad y conocimiento por descripción.*

Con el fin de profundizar en la disputa sobre la semántica de los términos singulares, atenderemos seguidamente a las tesis de Russell, quien, en cierto sentido, sigue los desarrollos de Frege y, en otro sentido, nos acerca a los partidarios de las Nuevas Teorías de la Referencia. La mejor manera de introducirnos en las ideas de este autor más relevantes para dicha disputa es acudir a (1905), un trabajo en el que Russell ofrece un análisis de aquellas oraciones en cuya forma gramatical figuran descripciones definidas. La idea central de Russell al respecto es que no resulta posible analizar una oración como “El autor de *El Quijote* es manco” en tanto que una mera atribución de una propiedad a un individuo, esto es, como si su forma lógica fuera “Fa”, siendo “F” el predicado que designa la propiedad de ser manco y “a” una constante de individuo que designa a Cervantes. Russell afirma que, de suscribir este análisis “ingenuo”, nos veríamos involucrados en una serie de dificultades, siendo dos de las más importantes aquellas que explicaremos a continuación.

En primer lugar, no parece fácil establecer el valor de verdad de aquellas oraciones en las que figura como sujeto una descripción definida carente de denotación. Por ejemplo, pensemos en la siguiente oración:

(7) El último chotacabras jamaicano vivo en la actualidad está en el zoológico de Madrid.

Teniendo en cuenta que esa clase de aves se extinguió hace más de treinta años, no podemos decir que esta oración sea verdadera, puesto que no hay ningún chotacabras jamaicano en el zoológico de Madrid. Pero para poder afirmar su falsedad parece que deberíamos localizar al último chotacabras jamaicano vivo en la actualidad en otro lugar distinto del aludido en (7). Sin embargo, por hipótesis, se trata de una empresa imposible.

La segunda dificultad suscitada por la interpretación ingenua viene planteada por la sustitución de descripciones definidas por otros términos con la misma referencia en contextos de atribución de creencias, una dificultad que, como puede verse, está estrechamente emparentada con las preocupaciones de Frege en (1892). Por ejemplo,

Rufino, completamente horrorizado, lee en el periódico que la policía está buscando a una misteriosa mujer que ha cometido múltiples homicidios en Vallecas, dado lo cual podríamos informar de sus pensamientos diciendo:

(8) Rufino cree que la famosa asesina de Vallecas es una persona poco recomendable.

Este hablante desconoce que Eva, su apreciada nuera, es el individuo que constituye la denotación de la descripción definida “la famosa asesina de Vallecas”. Sin embargo, si en nuestro informe sobre los pensamientos de Rufino sustituyéramos esa descripción definida por el nombre propio de la persona denotada por esa descripción, diríamos algo falso de él, a saber, que cree que Eva es una persona poco recomendable.

La solución de Russell pasa por cuestionar el supuesto en el que se sustentan estos dos puzzles, a saber, el supuesto de que la contribución de una descripción definida a la proposición expresada por las oraciones en cuya forma gramatical figura es el individuo denotado por dicha descripción. La lectura de este autor nos libera de la necesidad de localizar en alguna parte al ave denotada por “el último chotacabras jamaicano vivo en la actualidad”. Asimismo, la racionalidad de Rufino no se vería en entredicho al decir de él que cree y no cree de Eva que es una persona poco recomendable. La propuesta de Russell para analizar las descripciones definidas consiste en entender que su contribución semántica equivale a la de una estructura cuantificacional del tipo “existe un individuo y sólo un individuo que es F ”. Dicha estructura establece la condición de unicidad y existencia del individuo denotado por la descripción, por lo que si no hubiera ninguna entidad que satisfice dicha condición, esto es, que posee la propiedad expresada por ella (como en “el último chotacabras jamaicano vivo en la actualidad”), o si hay más de una entidad que la posee, entonces la descripción definida carecería de denotación, pero no de aportación semántica. Así, el valor semántico de una descripción definida no es el individuo, si lo hay, denotado por dicha descripción, sino una estructura cuantificacional que contiene la propiedad designada por el predicado presente en la descripción definida³⁵ que es abreviada mediante el nombre propio.³⁶

³⁵ Aunque una descripción puede contener varios predicados, que designan diversas propiedades, en lo siguiente, en aras de la brevedad, me atenderé a la forma de expresión recién usada.

³⁶ Hay que señalar que el dictamen de Russell en lo referente a la aportación proposicional de las descripciones definidas es aceptado por los partidarios de la Teoría de la Referencia Directa, al menos en lo que concierne a ciertos usos de las descripciones definidas. A este respecto conviene traer a colación la distinción entre los *usos atributivos* y los *usos referenciales* de las descripciones definidas propuesta por Donnellan en (1966). Este autor admite que el examen de Russell es adecuado para el primero de estos usos, mediante el cual pretendemos referirnos a quienquiera que sea el único individuo que satisfice la propiedad expresada por la descripción definida. Si, por ejemplo, Moisés dijera “el autor de este desastre

Para Russell este análisis de la contribución semántica de las descripciones definidas constituye la base de sus consideraciones sobre el valor semántico de la mayoría de los nombres propios.³⁷ En efecto, Russell afirma que los *nombres propios en el sentido usual del término*, nombres como “París”, “Raquel” o “Francisco”, tienen el mismo comportamiento semántico que las descripciones definidas. La mejor manera de entender esto es contrastar esta tesis con la afirmación de Russell de que los únicos nombres propios que tienen un comportamiento directamente referencial y cuyo valor semántico, en consecuencia, es idéntico al individuo que designan son los *nombres lógicamente propios*. Tales nombres son los demostrativos “esto”, “eso”, “aquello” en tanto que usados para referirnos a los datos sensoriales presentes a nuestra mente, datos que conocemos de manera directa. La clase de conocimiento que corresponde a los datos sensoriales (actualmente presentes a la mente o evocados en el recuerdo), a ciertos universales y al yo³⁸ es, según Russell, el *conocimiento por familiaridad* (“knowledge by acquaintance”). Conocemos por familiaridad aquello de lo que somos directamente conscientes, sin necesidad de procesos inferenciales o apoyo epistémico en otras verdades conocidas.³⁹ Sin embargo, el conocimiento de los objetos físicos y de las otras mentes corresponde a otra categoría epistemológica: el *conocimiento por descripción*. Este conocimiento se apoya en inferencias a partir de los datos sensoriales, conocidos por familiaridad. Por ejemplo, conozco por descripción la puerta de mi habitación como

lo pagará caro” al ver su valioso jarrón hecho añicos en el suelo sin saber quién ha sido el responsable de este accidente, estaría usando atributivamente la descripción “el autor de este desastre”. Al hacer un uso atributivo de la descripción definida “el autor de este desastre”, la aportación semántica de esta expresión es, como Russell afirma, una condición descriptiva que ha de cumplir el (único) individuo que posea la propiedad expresada por la descripción. Sin embargo, en ocasiones hacemos un uso distinto de las descripciones definidas, a saber, cuando tenemos la intención de hablar acerca de un individuo específico. Si Moisés hubiera dicho “el autor de ese desastre lo pagará caro” señalando a su hijo mayor habría hecho un uso referencial de la descripción definida involucrada en su preferencia. Mientras que en un uso atributivo de una descripción, el individuo *denotado* por ella es aquel individuo que, en efecto, posee la propiedad expresada por la descripción, esa misma descripción usada referencialmente se *refiere* a un individuo aunque éste no posea la propiedad en cuestión. Si, por ejemplo, Moisés se hubiera equivocado al suponer que su hijo mayor es responsable de la ruptura del jarrón, sin embargo, cabría decir, a partir de su conducta acusatoria, que se estaba refiriendo a él mediante el uso de la descripción “el autor de este desastre”. Los valedores de la Teoría de la Referencia Directa consideran que las descripciones definidas usadas referencialmente tienen un comportamiento semántico directamente referencial.

³⁷ Ésta es una tesis que no siempre mantuvo Russell. En una carta que dirige a Frege con fecha de 1904, Russell afirma que el monte “Mont Blanc es en sí mismo un componente de lo que de hecho se asevera en la proposición ‘Mont Blanc tiene más de 4000 metros de altura’”. Cf. Salmon/Soames (1988), pp. 56-57.

³⁸ No obstante, Russell cambió de opinión en lo que respecta a la clase de conocimiento que corresponde a nuestro propio yo. Aunque en obras tempranas como (1911), p. 134 y (1912), p. 27, sostuvo que conocemos nuestro yo por familiaridad, en (1914b), pp. 58-59, se muestra dubitativo al respecto. Según Chisholm, Russell ofreció su posición más extrema en (1959), p. 101, texto en el que llega a afirmar que el sujeto es una ficción lógica cuya introducción en el discurso no se fundamenta en observación alguna, cf. Chisholm (1974), pp. 212-216.

³⁹ Cf. Russell (1914), pp. 127 y ss., y Russell (1989), pp. 25 y ss.

aquello que causa tales y cuales datos sensoriales (cierta impresión de color, cierta sensación táctil). Según se aprecia, los nombres propios en el sentido usual del término (v. gr., “España”, “María”, “Hilary Putnam”) designan, a partir de este criterio, entidades que conocemos por descripción.

La razón de que Russell niegue un comportamiento semántico directamente referencial a los nombres propios en el sentido usual del término se fundamenta en consideraciones epistemológicas. Así, para este autor, *los constituyentes de un pensamiento, que es la proposición expresada por una oración, han de ser entidades conocidas por familiaridad* (Russell 1910, p. 23). Los nombres propios en el sentido usual del término designan objetos físicos, individuos y lugares, todos ellos objetos que conocemos por descripción. Si en la expresión de un pensamiento aparece un nombre propio en el sentido usual del término como “Francisco”, la entidad designada por este nombre propio no podría considerarse un constituyente del pensamiento, en la medida en que sólo conocemos dicho individuo por descripción o, lo que es igual, no podemos acceder epistémicamente a él de manera directa e inmediata atendiendo a los criterios de Russell. Más bien, lo que conocemos por familiaridad son los datos sensoriales particulares que, según inferimos, causa la entidad designada por “Francisco”, así como los universales que constituyen algunas de las propiedades que atribuimos a dicha entidad. La idea de Russell es que los nombres propios en el sentido usual del término no tienen un comportamiento directamente referencial porque son descripciones definidas abreviadas o encubiertas, es decir, cada uno de estos nombres abrevia aquella descripción definida que el hablante asocia con el nombre propio. En este sentido, cuando formo en mi mente un pensamiento al tomar en consideración una oración, los componentes de ese pensamiento han de ser, de acuerdo con Russell, entidades que conozco por familiaridad. Como vemos, no tengo esa clase de acceso epistémico con respecto a los individuos designados por nombres propios como “Francisco” o “Ana”, sino con respecto a los datos sensoriales y ciertos universales, debido a lo cual aquellos pensamientos en cuya expresión figura un nombre lógicamente propio y una propiedad, como “esto es azul” o “eso está caliente”, expresan, para Russell, proposiciones cuyos componentes son los datos sensoriales designados por “esto” o “eso” y los universales designados por “azul” o “caliente”. De este modo, los nombres lógicamente propios (y los términos generales que designan universales que conocemos por familiaridad) aportan sus *designata* a la proposición expresada por las oraciones en las que aparecen.

Repárese en que, según lo dicho, Russell liga el comportamiento semántico de las expresiones con el conocimiento que podemos obtener de los objetos designados por ellas. La contribución de un término a las proposiciones expresadas por las oraciones que lo contienen coincide con la contribución de dicha expresión al pensamiento que un hablante tiene en mente cuando toma en consideración dichas oraciones. En el caso de los nombres lógicamente propios, la proposición expresada por las oraciones en las que figuran estos términos ha de tener como constituyentes las entidades designadas por ellos y, paralelamente, el pensamiento que el sujeto se forma cuando considera estas oraciones ha de incluir como constituyentes a dichas entidades. Esto implica que si un nombre lógicamente propio careciera de designación, la oración en la que éste aparece carecería de significado y, por ello, el sujeto fracasaría en su tarea de formarse un pensamiento correspondiente al aceptar la oración. La razón de esto es que Russell defiende que no podemos tomar en consideración una suposición o juzgar la verdad de la oración que la expresa sin saber propiamente qué es aquello que estamos pensando o juzgando. En su caso “saber propiamente” se traduce, como se ha visto, en conocer por familiaridad los constituyentes de la proposición así considerada. Podemos entender, con ello, que el pensamiento que un sujeto se forma al considerar una oración en la que figura un nombre lógicamente propio es un pensamiento que versa esencialmente sobre la entidad designada por este término. Si este término no designa objeto alguno, el sujeto no tiene posibilidad de “saber propiamente” qué es aquello que se está tomando en consideración. Lo que nos interesa destacar de este análisis es el hecho de que esta clase de pensamientos *requiere* para existir de la existencia de la entidad designada por el nombre lógicamente propio. La denominación usual de este fenómeno supone catalogar a este tipo de pensamientos como *pensamientos objeto-dependientes*. Dado que remiten esencialmente a una entidad individual, de no existir dicha entidad, es decir, de no saber qué es aquello sobre lo que está pensando, el sujeto fracasaría a la hora de tomar en consideración un pensamiento correspondiente.

Lo dicho hasta el momento nos permite apreciar que la propuesta semántica de Russell concerniente a los nombres propios en el sentido usual del término conecta con la de Frege. Ambos consideran que el valor semántico de los nombres propios o, en el caso de Russell, de los nombres propios en el sentido usual del término *no* es el individuo designado por ellos. Para Frege la contribución de los nombres propios al pensamiento expresado por las oraciones en las que aparecen es su sentido, no su referente, y el sentido es, para cada nombre propio, equivalente al sentido de alguna

descripción definida. Para Russell, en cuanto que los nombres propios en el sentido usual del término sólo son descripciones definidas abreviadas, su valor semántico coincide con el valor semántico de las descripciones definidas que abrevian y, por tanto, su aportación proposicional es una estructura cuantificacional.

En el terreno psicológico, el paralelismo entre las ideas de estos autores también parece patente. Russell, por una parte, niega que el pensamiento que un sujeto se forma al juzgar una oración en la que figura un nombre propio en el sentido usual del término sea un pensamiento que contenga como constituyente el individuo designado por el nombre. Su contribución viene dada, más bien, por una estructura cuantificacional que contiene la propiedad designada por el predicado presente en la descripción definida que abrevia el nombre propio. Frege afirma, por otra parte, que el pensamiento que un hablante capta al entender una oración no es más que el sentido de dicha oración y, en el caso de los nombres propios, el sentido es una dimensión semántica distinta de su referente, por lo que dicho pensamiento tampoco incluiría, en su caso, el individuo designado por el nombre propio.

La Teoría Clásica de los términos singulares, cuyos representantes más destacados son Frege y Russell, es, como apuntamos al principio, *una teoría semántica de carácter internista*. El valor semántico de los nombres propios y descripciones es una entidad situada en el ámbito interno al sujeto, en el sentido de que los hablantes que usan dichas expresiones poseen un *pleno acceso epistémico* a la misma. Nótese que no establecemos con ello el carácter mental o puramente subjetivo del valor semántico. Frege se cuida mucho de no afirmar tal cosa de los sentidos y, sobre todo, de los pensamientos (1918, p. 45 y ss.). No obstante, tanto los sentidos como los pensamientos son entidades que “captamos” mediante nuestra actividad subjetiva de pensar y, en esta medida, resultan perfectamente accesibles al hablante que entiende una expresión. No en balde Frege establece una ecuación entre el sentido de un término lingüístico y el valor cognitivo que éste tiene para el sujeto. El tinte internista de las propuestas de Russell también es claro. Los constituyentes de una proposición, del contenido del pensamiento que un sujeto se forma al tomar en consideración una oración, han de ser todos ellos entidades que conocemos por familiaridad. Los universales, los datos sensoriales y el propio yo agotan el repertorio de entidades con respecto a las cuales el sujeto posee esta clase de acceso epistémico tan especial. Para determinar, desde estas teorías, si dos sujetos tienen o no la misma creencia al decir “Sócrates es calvo” debemos atender, bien al sentido de las expresiones involucradas en sus preferencias, bien a aquellas entidades de

carácter interno que constituyen los componentes de ese pensamiento. En cualquier caso, estableceremos la identidad del contenido proposicional creído por estos hablantes atendiendo a factores internos a ellos mismos, esto es, elementos presentes en sus respectivas vidas cognitivas.

1.4 *Teoría de la Referencia Directa*

Examinada la aproximación tradicional al comportamiento semántico de los nombres propios, damos paso a las críticas y propuestas de los partidarios de la Teoría de la Referencia Directa, quienes sostienen que los nombres propios ordinarios, es decir, los nombres propios en el sentido usual del término tienen un comportamiento semántico directamente referencial, similar al que Russell asignaba a los nombres lógicamente propios.⁴⁰ Para esta Teoría, la contribución semántica de estas expresiones es el individuo que ellas designan, y no un complejo descriptivo, como propone la Teoría Clásica internista. Esta afirmación viene respaldada por la intuición de que un oyente no entiende lo que expresa una oración que involucra un término directamente referencial a menos que pueda identificar la referencia de ese término, yendo más allá del posible contenido descriptivo del término (Recanati 1993, p. 38).

La estrategia de estos autores para aportar plausibilidad a su tesis comienza por rebajar las exigencias del estándar epistemológico que Russell pone en juego al hablar del conocimiento que podemos obtener de las entidades designadas por términos singulares directamente referenciales (en su caso, los nombres lógicamente propios). En particular, nos proponen deshacernos de la distinción entre conocimiento por familiaridad y conocimiento por descripción. De esta forma, si aceptamos un compromiso epistémico mucho menos estricto que el russelliano, parece más razonable calificar a los nombres propios (ordinarios o en el sentido usual del término) como términos directamente referenciales. El comportamiento directamente referencial de una expresión supone que su aportación a la proposición expresada por las oraciones en las que figura es el individuo designado por ella. Para entender esto, tomemos en consideración estas dos oraciones:

⁴⁰ La mayor parte de los defensores de la Teoría de la Referencia Directa también atribuyen un comportamiento directamente referencial a los términos índice y a las descripciones definidas usadas referencialmente. Sin embargo, no examinaremos con detalle lo referente a estos términos singulares.

(12) Bertín Osborne es un talentoso cantante.

(13) Norberto Ortíz Osborne López-Valdemoro Marengo es un talentoso cantante.

Según la Teoría de la Referencia Directa, (12) y (13) expresan una misma proposición, cuya verdad depende de que el individuo designado mediante estos nombres propios (uno y el mismo) posea la propiedad que ambas oraciones le adscriben. De acuerdo con estos autores, la proposición que expresan aquellas oraciones en las que figura un término singular directamente referencial es una *proposición singular*, esto es, un tuplo o secuencia entre cuyos componentes figura el referente de tal término singular. En el caso de (12) y (13) se trataría de una secuencia de dos elementos formada por el individuo designado por estos términos y la propiedad que se les atribuye (Kaplan 1989a, p. 483). Del mismo modo, la oración “Noelia protesta” expresa, para la Teoría de la Referencia Directa, una proposición singular formada por el individuo particular designado por el nombre propio “Noelia” y la propiedad que se le adscribe. Esta proposición puede describirse así: <Noelia, protestar>.⁴¹ Los partidarios de la Teoría de la Referencia Directa sostienen, con ello, que el valor semántico de los términos directamente referenciales equivale a la entidad que designan y, por tanto, al contrario que para Russell, la contribución proposicional de los nombres propios en el sentido usual del término – como “Noelia” – no involucra una estructura cuantificacional que contiene ciertas propiedades (aquellas que serían expresadas por la descripción definida que, presuntamente, el nombre abreviaría). Los defensores de la Teoría de la Referencia Directa rechazan la pretensión de Frege y Russell de identificar el valor semántico de los nombre propios (o de la mayoría de ellos, en el caso de Russell) con el valor semántico de ciertas descripciones definidas, concretamente de la descripción definida que, según Frege, nos da el sentido de un nombre propio o que, según Russell, el nombre propio abrevia.

Ahora bien, es digno de señalarse que la Teoría de la Referencia Directa (a diferencia del Millianismo)⁴² es una teoría sobre el valor semántico de ciertas expresiones singulares que no prejuzga la cuestión sobre los mecanismos de designación de dichas expresiones. Dicho de otra manera, los partidarios de esta teoría

⁴¹ Para un tratamiento de las proposiciones singulares como secuencias formadas por individuos y propiedades, véase Richard (1990), p. 110.

⁴² Consúltese al respecto la nota 47 del presente capítulo.

no rechazan necesariamente la presencia de patrones de propiedades en el vínculo referencial entre la expresión y lo designado por ella.⁴³

La extensión de la categoría “término directamente referencial” a los nombres propios tiene dos consecuencias importantes para nosotros. La primera de ellas es que la dependencia de objeto afecta en un principio a un mayor número de pensamientos, mientras que para Russell, sólo los pensamientos en cuya expresión figura un nombre lógicamente propio son objeto-dependientes.⁴⁴ Ello se debe a que se defiende que los nombres propios tienen un comportamiento semántico directamente referencial y que, consiguientemente, las proposiciones expresadas por las oraciones en las que figuran estos términos son de carácter singular, por lo que tenemos que admitir que los pensamientos que los sujetos se forman al tomar en consideración estas oraciones son, asimismo, singulares y, por ello, objeto-dependientes. Esto tiene un efecto adicional cuando examinamos expresiones lingüísticas en las que figura un término directamente referencial que carece de referente (típicamente, un nombre propio), a saber, que aquel hablante que emita una preferencia en la que figura un término directamente referencial vacío (i) no estará realizando una aserción genuina y (ii) el pensamiento que expresa será un pensamiento vacío de contenido. En suma, la preferencia empleada no expresará una proposición completa al albergar un término carente de contribución proposicional.⁴⁵

⁴³ En este sentido, la mayor parte de estos autores admiten la presencia de complejos predicativos de esa clase en la relación de designación que se establece entre los términos indéxicos, y aquellos individuos a los que se refieren. Kaplan, particularmente, distingue en (1978) dos componentes en el significado de los indéxicos, a saber, su *carácter* y su *contenido*. Mientras que el contenido de un término indéxico constituye su contribución a la proposición expresada por las oraciones en las que figura, su carácter es aquel componente del significado de uno de estos términos que establece el modo en que el contenido viene determinado por el contexto. Si, por ejemplo, María dice “hoy llueve” y mañana dice “ayer llovió” estas oraciones expresan la misma proposición en virtud del hecho de que el contenido de los indéxicos “hoy” y “ayer” en los contextos respectivos es el mismo. Sin embargo, el carácter de estos términos, la regla que determina su contenido dado el contexto de uso es una regla distinta. Para el término “hoy” bien podría ser “el día presente”, mientras que para “ayer” sería “el día inmediatamente anterior”. Como podemos ver, la referencialidad directa de los términos indéxicos no queda puesta en entredicho por este análisis. La aportación semántica de estos términos, su contenido, es el individuo que designan. El carácter de un término indéxico es “externo a la proposición expresada” (Recanati 1993, p. 30) y, por tanto, las propiedades y especificaciones generales involucradas en el carácter no forman parte de lo que expresa la oración en la que figura el término indéxico. Su papel se ciñe a la determinación contextual del contenido que corresponde al uso de un indéxico, esto es, a forjar el vínculo entre la expresión y lo designado por ella.

⁴⁴ Hay que decir, sin embargo, que autores como Evans y McDowell aceptan la existencia de pensamientos objeto-dependientes dentro de un enfoque neo-fregeano. Esta visión mixta analiza nuestros pensamientos singulares (pensamientos que versan esencialmente sobre un objeto) como pensamientos que involucran *sentidos objeto-dependientes*. Evans en (1982), pp. 7 y ss., defiende que esta visión no resulta ajena al propio Frege. Véase también McDowell (1984).

⁴⁵ Algunos partidarios de la Teoría de la Referencia Directa han elaborado diversas propuestas para explicar el comportamiento semántico de los nombres carentes de referente o nombres vacíos. Debido a

La segunda consecuencia de la extensión del comportamiento directamente referencial a la mayor parte de los nombres propios es la siguiente. Si en dos oraciones como “Pablo Neruda escribió poesía” y “Neftalí Ricardo Reyes Basoalto escribió poesía” es uno y el mismo individuo el que designan los nombres propios involucrados en ellas, entonces ambas expresarán, para la Teoría de la Referencia Directa, una y la misma proposición singular, visto que el mismo individuo, amén de la misma propiedad, figuran en la proposición expresada por esas oraciones. Ahora bien, de aceptar que la proposición expresada por ambas oraciones es idéntica al contenido de la creencia del hablante que las profiere, el defensor de la Teoría de la Referencia Directa tendría que ofrecer una solución alternativa al problema que Frege resolvió mediante la postulación de los sentidos, a saber, que un hablante puede aseverar una de las oraciones mencionadas y negar la verdad de la otra oración. Si, de acuerdo con la Teoría de la Referencia Directa, las creencias de este hablante tienen el mismo contenido porque éste es idéntico a una misma proposición singular, ¿no es un síntoma de irracionalidad afirmar y negar *lo mismo*? Los valedores de la Teoría de la Referencia directa han elaborado distintas propuestas para lidiar con este problema, siendo una de las más elaboradas la propuesta de Salmon, cuyas líneas fundamentales describiremos en lo que sigue. Este autor mantiene – en tanto que partidario de la Teoría de la Referencia Directa – que dos oraciones como “a es F” y “b es F”, siendo “a” y “b” términos correferenciales, expresan una y la misma proposición y, consiguientemente, cualquier hablante que asevere ambas oraciones tendrá una misma creencia. En segundo lugar, atendiendo a una situación en la que un agente asiente ante la proferencia de “a es F” y niega la verdad de “b es F”, Salmon afirma que dicho hablante mantendrá actitudes contrarias hacia un mismo contenido, pero esto no es una muestra de irracionalidad por su parte. La razón es que existen *guisas*, maneras de tomar el contenido, que ejercen la

que se trata de una cuestión tangente con respecto al tema que nos ocupa, sólo tomaremos en consideración las ideas contenidas en Kripke (2011a), un trabajo dedicado a resolver este problema en relación a los nombres propios de personajes de ficción, una clase de nombres vacíos, aun cuando sus conclusiones puedan ser extrapoladas a casos en los que un hablante usa un nombre propio vacío y, por error, considera que sí tiene referente en el mundo actual (Kripke 2011a, p. 61). De acuerdo con este autor, las oraciones que figuran en un relato de ficción no expresan genuinas proposiciones, sino “proposiciones supuestas” (“pretended propositions”). En lo que se refiere a la posibilidad de evaluar semánticamente esta clase de proposiciones, Kripke sostiene que muchas de ellas son susceptibles de recibir un valor de verdad. Así, para explicar la intuición de que la oración “Sherlock Holmes era un detective” es, en cierto sentido, una oración verdadera, este autor otorga un cierto estatus ontológico a los “referentes” de los nombres propios de personaje de ficción, declarando así que se trata de *entidades ficticias*. De este modo, podremos decir que la oración que atribuye la propiedad de ser un detective a la entidad ficticia designada por el nombre “Sherlock Holmes” es una oración verdadera porque en el contexto del relato de ficción relevante la entidad ficticia designada por ese nombre propio posee dicha propiedad.

función de intermediarios epistémicos entre el hablante y el contenido. Debido a esto, un mismo contenido (idéntico a la misma proposición singular que expresan “a es F” y “b es F”) puede presentarse con guisas distintas al mismo individuo y dar lugar a actitudes epistémicas contrarias por su parte (cf. Salmon 1986).⁴⁶

Llegado este punto, ofreceremos un breve resumen de lo expuesto en esta primera sección. Frente a las propuestas clásicas de Frege y Russell, la Teoría de la Referencia Directa pone en práctica un enfoque externista en el análisis del comportamiento semántico de los nombres propios. Esta Teoría defiende que el valor semántico de tales términos, su contribución proposicional, es el objeto designado por ellos, oponiéndose de esta manera a la identificación clásica del valor semántico de los nombres propios con complejos de propiedades (conocidas por el hablante e “internas” a él en este sentido) que pretenden ser poseídas únicamente por el objeto al que se refieren. Las proposiciones expresadas por las oraciones en las que figuran los nombres propios y, en general, los términos directamente referenciales son, para Teoría de la Referencia Directa, proposiciones que involucran como constituyente al individuo, generalmente externo al hablante, que estos términos designan. Por ello, los pensamientos que los hablantes expresan mediante ese tipo de oraciones se dirigen hacia un contenido proposicional singular y, en consecuencia, son actitudes psicológicas que versan esencialmente sobre la entidad a la que se refiere el término directamente referencial involucrado en la expresión del pensamiento.

⁴⁶ En general se pueden distinguir dos filiaciones entre los autores que rechazan la postulación de sentidos para resolver el puzzle de Frege. Por un lado, las propuestas neo-russellianas y, por otro, las propuestas contextualistas. Los autores que defienden el primer enfoque (entre los que se encuentran McKay, Salmon y Soames) mantienen que la proposición expresada por la cláusula-que de las atribuciones “S cree que a es F” y “S cree que b es F” es una y la misma, idéntica a la proposición singular formada por el par ordenado que incluye el individuo designado por “a” y “b” (el mismo, por hipótesis) y la propiedad designada por “F”. La postulación de guisas, en el caso de Salmon, no alteraría esta imagen, puesto que las distintas maneras de tomar el contenido son externas al propio contenido, idéntico en ambos casos. La consecuencia que se desprende de ello es que las cláusulas-que “a es F” y “b es F” tienen exactamente las mismas condiciones de verdad, a saber, que el individuo designado mediante “a” y “b” tiene la propiedad expresada por “F”. Esto significa que si un informe es verdadero el otro debe serlo también y, puesto que “S cree que a es F” es verdadero, “S cree que b es F” también lo será, aun cuando S rehúse aceptar la verdad de “b es F”. Las propuestas contextualistas, por su parte, admiten que la proposición expresada por “b es F” puede ser distinta de la expresada por “a es F” en función de los factores presentes en el contexto de la preferencia. Así, por ejemplo, Crimmins y Perry en (1989) mantienen que en nuestros informes mentales existe un componente indéxico oculto que selecciona, entre otras cosas, las ideas y nociones particulares que un hablante tiene en mente sobre los individuos y las propiedades externas acerca de los que versa el pensamiento en cuestión. De esta forma, como las ideas y nociones que un hablante asocia con el *designatum* de “a” y “b” son distintas en función de si es “a” o “b” su designador, cobra sentido que este hablante mantenga actitudes contrarias hacia las oraciones “a es F” y “b es F”. Dado que, además, esas ideas y nociones forman parte del contenido que dicho hablante expresa al proferir estas oraciones, los informes “S cree que a es F” y “S cree que b es F” tendrán distintas condiciones de verdad y, consiguientemente, si uno de ellos es verdadero no tiene por qué serlo el otro.

2. *Externismo semántico (II): Millianismo, Kripke y Putnam*

En la primera sección de este capítulo nos hemos ocupado de determinar el origen y sentido de la polémica entre internistas y externistas acerca del *valor semántico* de los nombres propios. Esta segunda sección estará dedicada al siguiente punto sobre el que discrepan ambas teorías semánticas, a saber, la naturaleza de la *conexión referencial* entre nuestras expresiones lingüísticas, especialmente los nombres propios y los términos de género natural, con las entidades que designan.

Antes de entrar a fondo en esta cuestión se hace preciso, sin embargo, realizar dos aclaraciones. En primer lugar, recordemos que aquellos autores que, encuadrados en las Nuevas Teorías de la Referencia, se oponen a la concepción internista de dicha conexión recibirán el nombre de Millianos.⁴⁷ En segundo lugar, debemos presentar la noción de término de género natural y atender a la forma en que la Teoría Clásica se hace cargo de la determinación de la referencia de esta clase de términos. Aunque no hay una caracterización precisa y unánimemente aceptada de la noción de término de género natural, asumiremos que los *términos de género natural* son una clase de términos generales que se requieren en nuestras explicaciones científicas,⁴⁸ como, por ejemplo, términos que designan elementos y compuestos químicos (“agua”, “oro”) términos que designan especies biológicas y botánicas (“tigre”, “haya”), así como términos de la teoría física (“electrón”).⁴⁹ En relación a las propuestas de la Teoría

⁴⁷ En este punto, es necesario introducir dos observaciones adicionales. En primer lugar, de acuerdo con lo indicado en la sección 1.1.1., emplearemos la denominación de “Millianismo” exclusivamente para aludir a aquellas propuestas que rechazan que la conexión referencial entre nuestras palabras y los aspectos del mundo que designan viene dada por ciertas descripciones que los usuarios del lenguaje han de tener en mente al usar esas palabras. De este modo, un autor milliano puede mantener que existe algún tipo de mediación entre nuestras expresiones y sus designaciones, sólo que dicha mediación no tiene un carácter descriptivo. En segundo lugar, debemos señalar que el hecho de ofrecer una denominación genérica (Nuevas Teorías de la Referencia) para agrupar las ideas de carácter externista y anti-fregeano en torno al valor semántico y a la determinación de la referencia no debería llevarnos a creer que esas Nuevas Teorías constituyen un bloque unitario de propuestas.

⁴⁸ Putnam caracteriza los géneros naturales como “clases de cosas que consideramos de importancia explicativa; clases cuyas características distintivas normales son ‘mantenidas juntas’ o incluso explicadas por mecanismos subyacentes” (Putnam 1970, p. 139).

⁴⁹ En este trabajo nos centraremos especialmente en los términos de género natural que designan sustancias químicas, debido a que la mayor parte del debate sobre la posibilidad de compatibilizar el externismo y el acceso privilegiado toma en consideración el escenario descrito por Putnam en su experimento mental en torno al uso del término de género natural “agua” (que designa una sustancia química), del que nos ocuparemos en la sección 2.2.2 de este capítulo. Por otra parte, es cierto que este autor examina, aunque con menos detenimiento, otros tipos de términos generales, a saber, los términos socio-legales, como “pediatra”, y los términos de género artificial. La extensión de esta última clase de términos está constituida por artefactos y objetos elaborados por la mano del hombre. “Lápiz” o “mesa” serían términos de esta clase, cuya extensión es, respectivamente, el conjunto de los lápices y el conjunto de las mesas. En lo que sigue, atenderemos sólo a las propuestas de Putnam relativas a los términos de

Clásica sobre la referencia de los términos de género natural, se propone, de nuevo, situar los elementos que explican el vínculo referencial en el ámbito cognoscitivo (“interno”) de los usuarios del lenguaje. Así, de acuerdo con esta teoría, el sentido de un término como “oro” vendría dado por una descripción que expresa ciertas propiedades que determinan la extensión del término y que, en ese caso, podría tratarse de la descripción “el metal brillante, maleable, amarillo y valioso”. El carácter interno de la determinación Clásica de la referencia es claro, tanto en el caso de los nombres propios como en relación a los términos de género natural. De este modo, en la medida en que un hablante posee competencia en el uso, bien de un nombre propio, bien de un término de género natural, tendrá, asimismo, un conocimiento (interno) de la descripción que o bien abrevia el nombre/que nos da su sentido, o bien que expresa las propiedades que típicamente poseen las entidades a las que se aplica el término de género natural. En este último caso, un hablante que usa el término “oro” podrá referirse con éxito al oro o, en sentido estricto, a porciones de oro, en virtud de su conocimiento de una descripción de esa índole.

Hechas estas aclaraciones, conviene enunciar aquí el contenido de las dos partes de esta segunda sección. En la primera subsección presentaremos las críticas desarrolladas por Kripke (1980) contra la Teoría Clásica de la Referencia de Frege y Russell, tanto en lo que respecta a los nombres propios como a los términos de género natural. En la segunda subsección abordaremos, en cambio, las ideas que Putnam desarrolla en (1975c) acerca de la semántica de los términos de género natural y repararemos especialmente en el conocido experimento mental de la Tierra Gemela, que será de gran relevancia para nuestro trabajo.

2.1 *Las críticas de Kripke a la Teoría Clásica de los nombres propios y de los términos de género natural. Sus propuestas al respecto*

En esta subsección tendremos ocasión de entender de qué manera Kripke se hace cargo de las insuficiencias de la Teoría Clásica de la Referencia de los nombres propios y de cierto tipo de términos generales, entre los que destacan los términos de género natural. Presentaremos, además, su propuesta sobre las propiedades semánticas de ambas clases de expresiones. El *locus classicus* sobre esta cuestión lo constituye el trabajo de Kripke

género natural por ser ésta la clase de expresiones involucradas en los experimentos mentales que este autor desarrolla en (1975c).

El nombrar y la necesidad. En lo que sigue atenderemos primeramente a los tres tipos de argumentos anti-descriptivistas descritos en esta obra en relación a los nombres propios para explicar, a continuación, cómo se extienden estas ideas a los términos de género natural.

El primero de los argumentos kripkeanos contra las pautas descriptivistas para la determinación de la referencia de los nombres propios es conocido como *argumento semántico* (Salmon 2005a, pp. 29 y ss.). Su punto de partida viene dado por la observación de Kripke de que en muchas ocasiones un hablante puede referirse con éxito a un individuo N aunque asocie con el nombre propio “N” una descripción definida que expresa una propiedad que N no posee. En un caso extremo, el acto referencial sería exitoso aunque ese hablante no tuviera en mente ninguna descripción singularizadora como la que exige la Teoría Clásica en ese sentido. Para aportar plausibilidad a esta idea, veamos un ejemplo presentado por el propio Kripke en (1980, p. 81), quien indica que la única información que asocian la mayoría de los hablantes con Cicerón es la de que este individuo era *un famoso orador romano*. Sin embargo, esta descripción (indefinida) no constituye la clase de descripciones (definidas y singularizadoras) que, según la Teoría Clásica, se precisan para que un hablante pueda usar con éxito un nombre propio (el nombre propio “Cicerón”, en este caso).⁵⁰

El segundo de los argumentos de Kripke se conoce como *argumento modal* y pretende mostrar que los nombres propios y términos de género natural, por un lado, y la mayor parte de las descripciones definidas, por otro, tienen comportamientos semánticos diferentes en contextos modales. Para explicar de manera apropiada este argumento es conveniente introducir la noción de mundo posible, una noción fundamental en la semántica kripkeana. Un *mundo posible* es, para este autor, un modo en que podemos concebir (estipular) que el mundo fuera diferente de cómo de hecho es. Concebir un mundo posible equivale a establecer una variación sobre el mundo actual,

⁵⁰ Donnellan, por su parte, ofrece en (1970) un segundo argumento para reforzar esta objeción a la Teoría Clásica partiendo de una asunción bastante plausible, a saber, que entre las descripciones definidas que la mayor parte de los hablantes asocia con el nombre propio “Tales” se encuentra la siguiente: “el filósofo griego que dijo que todo es agua”. Imaginemos que el autor de esa afirmación hubiese sido, en realidad, un oscuro eremita-filósofo griego con respecto al cual carecemos de conexiones históricas y, adicionalmente, supongamos que Tales nunca sostuvo ninguna de las tesis que se le imputan. Si ese fuera el caso, la teoría de la referencia de Frege y Russell predice que cuando usamos el nombre “Tales” nos estaríamos refiriendo, en realidad, a ese oscuro eremita, puesto que él es quien satisface las propiedades que asociamos con ese nombre. Sin embargo, este resultado es, en opinión de Donnellan, altamente contraintuitivo. Su idea es que aun cuando las descripciones definidas que asociamos con un determinado nombre propio no sean verdaderas del individuo designado por el nombre (algo que, según Kripke, es harto frecuente), este error no impide que nos estemos refiriendo a este individuo.

que por definición es un mundo posible, estipulando una situación alternativa o contrafáctica. Dicho esto, veremos de qué modo se desarrolla el argumento modal de Kripke.

En primer lugar, tengamos en cuenta que aquellos enunciados que establecen la identidad entre el nombre propio y la descripción definida que, según la Teoría Clásica, nos proporciona el sentido o significado de aquél (“N es el F”) habrán de ser enunciados analíticos, esto es, enunciados verdaderos en función de su significado. De acuerdo con la comprensión usual de la analiticidad, un enunciado analítico es, además de un enunciado conocido *a priori*, un enunciado necesario, lo que, empleando la terminología de Kripke, se traduce en que ha de ser un enunciado verdadero en todos los mundos posibles. Sin embargo, la mayoría de los enunciados del tipo “N es el F” no son enunciados necesariamente verdaderos.⁵¹ Así, podemos concebir un mundo posible en el que Aristóteles muere poco después de nacer. Al describir esta situación contrafáctica, advertimos que el nombre propio “Aristóteles” designa el mismo individuo que en el mundo actual fue discípulo de Platón y maestro de Alejandro Magno. No obstante, es obvio que en ese mundo posible la descripción definida “el discípulo de Platón y maestro de Alejandro Magno” no designa el mismo individuo que en el mundo actual, Aristóteles, puesto que, por hipótesis, en ese mundo tuvo una muerte temprana. Por esta razón, el enunciado “Aristóteles es el discípulo de Platón y el maestro de Alejandro Magno”, que es verdadero en el mundo actual, no sería verdadero en el mundo posible descrito y, en consecuencia, no sería un enunciado necesariamente verdadero. Kripke recoge estas consideraciones mediante la siguiente afirmación, a saber, que mientras que los nombres propios (términos singulares semánticamente simples) son *designadores rígidos*, la mayoría de las descripciones definidas, que son términos singulares semánticamente compuestos, no son designadores rígidos. A fin de comprender adecuadamente esta afirmación, es preceptivo presentar la noción kripkeana de designador rígido. Debido a que existe cierta ambigüedad en el uso de dicha noción por parte de Kripke en (1980) y (1971), nos serviremos de la caracterización que este autor ofreció a Kaplan cuando éste le solicitó una aclaración en este sentido. De este modo, dice Kripke, “un designador *d* de un objeto *x* es rígido si designa *x* con respecto a todos los mundos posibles en que *x* existe y nunca designa un objeto distinto de *x* con respecto a cualquier otro mundo posible” (Kaplan 1989b, p. 569). Como puede

⁵¹ Sí serían necesariamente verdaderos aquellos enunciados del tipo “N es el F” donde el predicado “F” expresara una propiedad necesaria o esencial de N.

apreciarse, esta definición deja abiertas dos posibilidades. La primera posibilidad es que un designador rígido designe el mismo objeto con respecto a todos los mundos posibles, mientras que la segunda posibilidad consiste en que designe el mismo objeto con respecto a todos los mundos posibles en los que dicho objeto existe y que carezca de referencia con respecto a aquellos mundos en los que el objeto no existe. Recurriendo a la terminología propuesta por Salmon en (2005a), podemos emplear una denominación diferente para cada una de estas dos posibilidades. De este modo, mientras que los *designadores rígidos obstinados* designan el mismo individuo en todos los mundos posibles, tanto si ese individuo existe en ese mundo como si no, los *designadores rígidos persistentes* designan el mismo individuo en todos los mundos posibles en que ese individuo existe y carecen de referencia en aquellos mundos posibles en los que el individuo no existe.

Por otro lado, en el Prefacio a la versión ampliada de (1980) el propio Kripke estableció una segunda distinción en el seno de los designadores rígidos, a saber, la distinción entre *designadores rígidos de iure* y *designadores rígidos de facto*. Un designador rígido lo es *de iure* si al fijar su referencia se *estipula* que su referente es el mismo independientemente de si estamos refiriéndonos al mundo actual o a otro mundo posible distinto del actual. Por su parte, en el caso de un designador rígido *de facto* el predicado contenido en la descripción se aplica al mismo objeto en cada mundo posible. Sin embargo, es conveniente debilitar la caracterización de rigidez *de facto* que acabamos de ofrecer con el fin de posibilitar las dos opciones siguientes, a saber, o bien que el predicado o término general con el que la descripción se ha formado se aplique al mismo objeto con respecto a todo mundo posible, o bien que se aplique al mismo objeto pero sólo en aquellos mundos posibles en los que el objeto exista y que carezca de referencia con respecto a todo otro mundo posible. Al permitir ambas posibilidades, los designadores rígidos *de facto* pueden ser o bien designadores (rígidos) obstinados, o bien designadores (rígidos) persistentes (vid. Fernández Moreno 2012, p. 112). Cabe decir, por último, que aunque Kripke no alude de manera expresa a la distinción entre designadores persistentes y obstinados, este autor vincula la noción de rigidez *de iure* con la de rigidez obstinada, en virtud de lo cual los designadores rígidos *de iure* serían designadores obstinados.⁵²

⁵² “Puesto que los nombres son rígidos *de iure* [...] digo que un nombre propio designa rígidamente su referente incluso cuando hablamos de [...] [mundos posibles] en los que ese referente no hubiera existido” (Kripke 1980, 21, n. 21).

Establecida esta clasificación, podemos completar la tesis central de la semántica de los nombres propios de Kripke diciendo que los nombres propios son designadores rígidos *de iure* (1980, p. 21, n. 21) mientras que la mayoría de las descripciones definidas no son designadores rígidos y las que designan rígidamente lo hacen *de facto*. Con el fin de aportar claridad sobre esta tesis es conveniente retomar el ejemplo anterior. Por una parte, el nombre propio “Aristóteles” es un designador rígido *de iure* mientras que la descripción definida “el maestro de Alejandro Magno” no es, por su parte, un designador rígido. Esta última afirmación se explica porque hay otros mundos posibles en los que un individuo distinto de Aristóteles se encarga de la educación de Alejandro Magno, lo cual supone que la descripción “el maestro de Alejandro Magno” designaría, en esos mundos, a estos individuos distintos del propio Aristóteles. No obstante, aunque esta descripción, como la mayoría de las descripciones definidas, no designe de manera rígida a su referente, Kripke admite que algunas descripciones sí lo hacen, aunque debemos recordar que tales descripciones no serían designadores rígidos *de iure*, sino designadores rígidos *de facto*.

Llegado este punto, presentaremos el último tipo de argumentos que Kripke elabora contra la Teoría Clásica de la Referencia, a saber, los *argumentos epistemológicos*. Estos presentan el mismo punto de partida que los argumentos modales, a saber, que aquellos enunciados de identidad que vinculan un nombre propio y la descripción definida que, según la Teoría Clásica, nos da su significado o sentido serán enunciados analíticos. Atendiendo a la comprensión más usual de la analiticidad, los enunciados analíticos son, además de necesarios, enunciados cuya verdad puede establecerse *a priori*. Sin embargo, si retomamos el ejemplo anterior, apreciamos que la verdad del enunciado “Aristóteles es el discípulo de Platón y el maestro de Alejandro Magno” no cabe establecerla *a priori*. Por esta razón, aunque la Teoría Clásica, según Kripke, está comprometida con la tesis de que este enunciado es analítico, este resultado atenta contra nuestra intuición de que su verdad no puede establecerse puramente *a priori*.

Una vez expuestos los tres tipos de argumentos que Kripke despliega contra la Teoría Clásica de la Referencia en relación a los nombres propios, en lo que sigue explicaremos la forma en que este autor extiende sus consideraciones anti-descriptivistas al ámbito de los términos de género natural, una extensión posibilitada por su convencimiento de que ambas clases de expresiones guardan un “estrecho parentesco” (Kripke 1980, p. 134). Para ello, recordemos que en las páginas precedentes

definimos los términos de género natural como términos generales requeridos en nuestras explicaciones científicas. Aunque el estatus ontológico de los géneros naturales y la extensión de la clase de los términos de género natural son cuestiones controvertidas, asumiremos provisionalmente esta breve caracterización, ya que será suficiente para exponer las bases de la semántica kripkeana de los términos de género natural. Con este fin, vamos a comenzar exponiendo las tres tesis centrales de Kripke sobre los géneros naturales, que, como veremos, tendrán una clara correspondencia con las ideas de Putnam sobre el mismo asunto. En primer lugar, para este autor cabe encontrar entre los miembros de un género natural un subconjunto de entidades a las que podemos considerar *miembros paradigmáticos* del género. En segundo lugar, este estatus de miembros paradigmáticos se deriva del hecho de que dichos ejemplares poseen una serie de *propiedades superficiales*, es decir, relativas a su apariencia externa, que consideramos típicas de los ejemplares del género en cuestión. En tercer y último lugar, presumimos que los ejemplares de un género natural comparten, en cuanto tales, una misma naturaleza o estructura interna o, lo que es igual, poseen *propiedades estructurales*, que son causalmente responsables de las propiedades superficiales.

Tras enunciar las pautas de la concepción kripkeana de los géneros naturales, podemos analizar sus consideraciones sobre la semántica de los términos de género natural. La tesis fundamental a este respecto es que existen una serie de similitudes en la semántica de los nombres propios y de los términos de género natural. A continuación, enumeraremos aquellos aspectos comunes en ambas clases de expresiones para examinar después de qué modo Kripke los justifica empleando un ejemplo del propio autor.

En primer lugar, este autor considera que tanto los nombres propios como los términos de género natural son *designadores rígidos*⁵³ y, en segundo lugar, dado el

⁵³ No obstante, es preciso realizar dos puntualizaciones. En primer lugar, si en la asignación de un carácter rígido para el modo de designación que corresponde a los términos de género natural, el término “referencia” fuera entendido, estrictamente, como el *conjunto* de individuos a los que el término se aplica, *ningún* término de género natural sería rígido, en la medida en que bastaría con que en un mundo posible la extensión del género contuviese un individuo menos que en el mundo actual para que el término de género natural relevante tuviera una extensión distinta en ese mundo posible, lo que impediría mantener idéntica su referencia entre mundos posibles. Por este motivo, autores como Martí/Fernández (2010) y Salmon han mantenido que los términos de género natural designan universales. Esta afirmación encuentra apoyo en aquellas observaciones de Kripke (1980, p. 136) en las que este autor concibe la relación entre un género natural y las entidades del género como una relación de *instanciación*. En adelante, suscribiremos la posición de estos autores por mor del argumento. En segundo lugar, dejamos abierta la posibilidad de que los términos de género natural puedan ser designadores obstinados o persistentes. En el caso de que fueran designadores obstinados, tales términos designarían la misma sustancia en todos los mundos posibles, tanto si esa sustancia existe en ese mundo como si no. En el caso

carácter rígido de la designación tanto de los nombres propios como de los términos de género natural, ambas clases de expresiones figuran en enunciados de identidad que, si son verdaderos, son necesariamente verdaderos, si bien la verdad de dichos enunciados será conocida *a posteriori*. Para explicar estas tesis tomaremos a modo de ejemplo uno de los enunciados de identidad formados por términos de género natural – enunciados que Kripke denomina “identificaciones teóricas” o “identidades teóricas” –, a saber, “el agua es H₂O”.⁵⁴ Según hemos dicho, si este enunciado es verdadero, se tratará de una verdad necesaria y, además, conocida *a posteriori*. De acuerdo con Kripke, el carácter necesario de dicho enunciado es consecuencia de que los dos términos involucrados en el enunciado de identidad son designadores rígidos.⁵⁵ De este modo, a partir de la rigidez del término de género natural “agua”, podemos afirmar que este término designará la misma sustancia que designa en el mundo actual en todo mundo posible, o al menos en todo mundo posible en el que exista el agua. Igualmente, la rigidez de la expresión “H₂O” nos permite afirmar que dicho término designará el H₂O, su designación en el mundo actual, en todo otro mundo posible o al menos en todo mundo posible en el que exista el H₂O. Por lo que respecta al carácter *a posteriori* del conocimiento que corresponde a dicho enunciado, puede comprobarse que dicho carácter es consecuencia de que se han requerido indagaciones de carácter empírico para

de que fueran designadores persistentes, los términos de género natural designarían la misma sustancia en todos los mundos posibles en que esa sustancia exista y carecerían de referencia en aquellos mundos posibles en los que la sustancia no exista.

⁵⁴ Kripke proporciona otros ejemplos de identidades teóricas: “el oro es el elemento químico cuyo número atómico es 79” (1980, p. 123), “el rayo es una descarga eléctrica” (1980, p. 132) y “el calor es movimiento molecular” (1980, p. 99). No obstante, “rayo” y “calor” son términos generales que designan fenómenos naturales más bien que términos de género natural. Por otro lado, aunque “H₂O” sea una abreviatura de una descripción definida (“el elemento químico cuyo número atómico es 79”) y no un término general, que es la clase de expresiones a la que pertenecen los términos de género natural, podemos obtener a partir de esta descripción un término general eliminando el artículo definido relevante, a saber, “sustancia cada una de cuyas instancias está compuesta de moléculas formadas por dos átomos de hidrógeno y un átomo de oxígeno”.

⁵⁵ “Las identidades teóricas, de acuerdo con la concepción que propongo, generalmente son identidades que involucran dos designadores rígidos y, por tanto, son ejemplos de enunciados necesarios *a posteriori*” (Kripke 1980, p. 140). Sin embargo, aunque aceptemos, por mor de la exposición, la tesis de Kripke de que el carácter necesario de las identidades teóricas es consecuencia de que los términos de género natural involucrados en ellas son designadores rígidos, hay autores que han puesto en cuestión esta afirmación, cf. Fernández Moreno (2012). Debe señalarse, además, que el ejemplo proporcionado por Kripke, “el agua es H₂O”, no involucra términos de género natural que designen entidades necesariamente existentes y, puesto que estamos dejando abierta la posibilidad de que los términos de género natural sean designadores persistentes – vid. nota 53 –, en aquellos mundos posibles en los que no existan instancias de H₂O podría ocurrir que la oración “el agua es H₂O” fuera falsa o carente de valor de verdad. En caso de que la oración en cuestión fuese falsa en dichos mundos posibles, con ella podemos formar un enunciado condicional de la forma “si existe el H₂O, entonces el agua es H₂O”, que será un enunciado necesariamente verdadero, es decir, verdadero con respecto a todos los mundos posibles.

establecer que el agua está compuesta de moléculas formadas por dos átomos de hidrógeno y uno de oxígeno.

De lo establecido por estas tesis sobre las identidades teóricas podemos extraer dos consecuencias. En primer lugar, si tomamos en consideración un mundo posible en el que hay porciones de un líquido cuya composición química es H_2O , entonces estamos tomando en consideración un mundo posible en el que *hay agua*. Más aún, Kripke afirma que si las porciones de H_2O de un mundo posible distinto del actual carecieran de *todas* las propiedades superficiales que poseen típicamente las porciones actuales de agua (ausencia de color, olor, sabor, etc.) ello no impediría que fueran porciones de agua. De este modo, la posesión de la propiedad estructural relevante por parte de una entidad constituirá una *condición suficiente* para que la entidad sea una instancia del género en cuestión. En segundo lugar, si en un mundo posible hay un líquido que comparte con el agua todas las propiedades superficiales pero que, al mismo tiempo, posee una composición química distinta de H_2O , ese líquido no será agua, sino una sustancia esencialmente diferente,⁵⁶ dado lo cual la posesión de la propiedad estructural por parte de una entidad no sólo será una condición suficiente de pertenencia al género, sino también una *condición necesaria*. Kripke sostiene, en consecuencia, que la propiedad estructural de un género natural constituye la *propiedad esencial* de dicho género, y, en el caso que nos ocupa, la propiedad esencial del agua – lo que determina qué es ser agua en todo mundo posible – es tener la composición química H_2O .⁵⁷

El último aspecto de las tesis de Kripke sobre la semántica de los términos de género natural que nos interesa destacar es que, de acuerdo con este autor, los enunciados en los que se atribuye a un género natural alguna de las propiedades superficiales que poseen típicamente los ejemplares del género son enunciados que expresan proposiciones contingentes. El motivo es que podemos concebir mundos posibles en los cuales los ejemplares del género (entidades que poseen en ese mundo posible la misma propiedad estructural que los ejemplares actuales del género) carecen de alguna de las propiedades superficiales relevantes. De la misma manera, Kripke también cuestiona el carácter *a priori* de estos enunciados. A modo de justificación, nos invita a concebir situaciones contrafácticas en las que un descubrimiento científico nos

⁵⁶ “Así como algo puede tener todas las propiedades mediante las cuales identificamos a los tigres en un principio sin que sea un tigre, de la misma manera podemos encontrar tigres que no tengan *ninguna* de las propiedades mediante las cuales los identificamos en un principio” (Kripke 1980, p. 121).

⁵⁷ Entendemos las propiedades esenciales como lo hace Kripke, a saber, en tanto que una propiedad que una entidad ha de tener en todos los mundos posibles en los que la entidad exista, exceptuando propiedades como la existencia y propiedades “triviales” como la identidad consigo misma.

lleva a concluir que los ejemplares de un género K no poseen, en realidad, la propiedad superficial relevante y que los hablantes han estado equivocados al asociar dicha propiedad con los ejemplares de K. De acuerdo con Kripke, el hecho de que descubrimientos científicos puedan hacer falsos estos enunciados pone en evidencia el carácter *a posteriori* de nuestro conocimiento de los mismos.⁵⁸

Las consideraciones semánticas precedentes le servirán a Kripke para combatir aquellas tesis que para él constituyen el eje de la Teoría Clásica del Significado y la Referencia de los términos de género natural. Dichas tesis son las siguientes: (i) el significado de un término de género natural viene dado por aquellas descripciones (o la mayoría de ellas) que los hablantes asocian con los ejemplares del género natural y (ii) la extensión de un término de género natural viene determinada por tales descripciones, de manera que un término de género natural designa el conjunto de entidades que poseen las propiedades que expresan las descripciones asociadas (Kripke 1980, p. 127) o un número suficiente de ellas (Kripke 1980, p.135).⁵⁹ La Teoría Clásica de las

⁵⁸ Para comprender esto, conviene establecer un contraste con enunciados cuyo carácter *a priori* es menos controvertido tales como “Si Juan es soltero, entonces es un hombre no casado”. Parece claro que no cabe pensar en ninguna experiencia posible tal que nos lleve a establecer la falsedad de este enunciado, es decir, que nos lleve a concluir que Juan es soltero y, a su vez, un hombre casado. La falsedad de este enunciado constituye una contradicción patente y su carácter de tal es perfectamente conocido en virtud del significado de las expresiones “soltero” y “hombre no casado”. Kripke, por tanto, diseña el siguiente test para establecer el carácter *a priori* de un enunciado, a saber, si cabe concebir alguna experiencia posible (como un descubrimiento científico empírico) tal que determine una variación en el valor de verdad de un enunciado, entonces dicho enunciado no tiene un carácter *a priori*.

⁵⁹ Hemos elegido la formulación más sofisticada de la Teoría Clásica sobre la semántica de los términos de género natural que Kripke nos proporciona en (1980, p. 127), un fragmento que contiene una breve recapitulación de estas tesis y de su evolución a manos de algunos de sus valedores. De este modo, Mill propone que los términos generales, incluidos los términos de género natural, vienen definidos por una conjunción de ciertas propiedades, que constituyen la connotación del término, la cual sirve, a su vez, para determinar su extensión. Asimismo, Kripke sitúa a Frege en la estela de Mill y afirma que este autor reservó al sentido de los términos generales el mismo papel (significado y elemento determinante de la extensión) que Mill atribuía a la connotación. No obstante, conviene aclarar que, en realidad, ninguno de estos autores formuló expresamente una teoría sobre la semántica de los términos de género natural en particular. En el caso de Russell, en cambio, sí encontramos una serie de consideraciones al respecto en (1948), las más importantes de las cuales son (i) un género natural es un conjunto de objetos que comparten una serie de propiedades conocidas, aun cuando la razón por la que estas propiedades (o la mayoría de ellas) coexisten en los miembros del género nos resulte desconocida (1948, p. 335); (ii) un término de género natural viene definido por propiedades conocidas (superficiales, en la terminología de Kripke) que, además, determinan su extensión, sin que se requiera que cada uno de los ejemplares del género posea *todas* esas propiedades (1948, p. 460) y (iii) aunque los géneros naturales (y los términos que los designan) resultan útiles para llevar a cabo inducciones pre-científicas, el conocimiento científico de aquellas propiedades que dan razón de la coexistencia de las propiedades conocidas hará prescindible la apelación a los géneros naturales (1948, p. 462). En lo que se refiere a los valedores de la Teoría Clásica del Significado y la Referencia de los términos de género natural posteriores a Frege y Russell, Kripke indica que su teoría al respecto resulta un tanto más sofisticada. En este sentido se defiende que aquello que nos proporciona el significado de un término general (y que determina su extensión) no es una única descripción o un único sentido, sino más bien un “racimo” de descripciones, de manera que el significado de esta clase de términos viene dado por *un número suficiente* de estas propiedades y su

propiedades semánticas de los términos de género natural tiene una consecuencia que entra en abierta contradicción con las propuestas kripkeanas, a saber, que para dicha Teoría los enunciados en los que se establece la identidad de un género natural con las propiedades que los hablantes asocian con el género son *definiciones* del género natural relevante, en el sentido de que proporcionarían el significado del término.⁶⁰ Si para la Teoría Clásica tales oraciones proporcionan el significado del término, entonces aquellos enunciados en los que se establece la identidad del término de género natural con aquellos predicados que los hablantes asocian con dicho término serán oraciones verdaderas en virtud del significado de sus expresiones componentes, es decir, oraciones *analíticas* que, en cuanto tales, expresan verdades necesarias conocidas *a priori*. Ahora bien, para comprender la crítica de Kripke a esta consecuencia de la Teoría Clásica, es conveniente tener en cuenta el siguiente hecho, a saber, que las propiedades que los hablantes asocian con los ejemplares de un género natural son, en su mayor parte, propiedades superficiales de dichos ejemplares. Esto supone que las oraciones analíticas que, de acuerdo con la Teoría Clásica, nos dan el significado de los términos de género natural son oraciones en las que se afirma la identidad del género con algunas de sus propiedades superficiales. Sin embargo, ya hemos visto que, para Kripke, no conocemos *a priori* la verdad de las oraciones de esta clase, puesto que su valor de verdad (o un cambio en el mismo) puede depender de posibles descubrimientos empíricos. Además, este tipo de oraciones tampoco expresan verdades necesarias, ya que hay mundos posibles en los cuales los ejemplares del género (entidades que poseen la propiedad estructural relevante) carecen de todas las propiedades superficiales que pueden figurar en el predicado de estas oraciones. Si cabe cuestionar el carácter necesario y *a priori* de aquellos enunciados que, según la Teoría Clásica, nos proporcionan el significado de los términos de género natural, entonces tales enunciados ni son analíticos ni, en consecuencia, las descripciones asociadas con el término de género natural constituyen su significado.

Tras haber dado cuenta de las críticas de Kripke a la Teoría Clásica de la Referencia de los nombres propios y de los términos de género natural, debemos

extensión queda constituida por el conjunto de individuos que poseen *un número suficiente* de tales propiedades.

⁶⁰ Hay que señalar, no obstante, que Kripke distingue dos sentidos en los que una descripción puede ser una definición de un designador, a saber, en tanto que dicha descripción fija la referencia de éste y en tanto que nos proporciona un sinónimo del designador relevante (1980, p. 59). Al decir que para la Teoría Clásica las descripciones que determinan la referencia de un término de género natural nos proporcionan una definición del género, hemos entendido el término “definición” en este segundo sentido.

preguntarnos qué explicación alternativa nos ofrece este autor para hacerse cargo de la naturaleza del vínculo referencial. O, dicho de otra manera, si la referencia de los nombres propios y de los términos de género natural que estamos tomando en consideración no requiere de descripciones definidas mediadoras, ¿en virtud de qué estas expresiones se refieren a sus *designata*? Kripke nos propone un “cuadro mejor” para responder a esta pregunta (1980, p. 93) cuyas pautas principales recogemos a continuación. En primer lugar, con respecto a los nombres propios, estas expresiones son introducidas en el lenguaje a partir de un *bautismo inicial*, formal o informal, en el que se fija su referencia mediante ostensión (“esto es N”) o descripción (“N es el F”). Es necesario señalar que, a pesar de que existan descripciones involucradas en el proceso de fijación de la referencia de un nombre, dichas descripciones (i) no son sinónimas del término en cuestión y (ii) no es necesario que los hablantes presentes en el bautismo las recuerden para que puedan usar exitosamente el término con posterioridad. Es más, tras el bautismo inicial, la expresión se transmite de los hablantes presentes en el acto a otros hablantes que estuvieron ausentes, los cuales no tienen por qué conocer tampoco aquellas descripciones involucradas en el bautismo inicial para poder usar el término con éxito.

En segundo lugar, con respecto a los términos de género natural el proceso de fijación de su referencia transcurre de manera muy parecida. Los hablantes introducen el término para designar un conjunto de entidades que, según se presume, comparten una misma estructura. Por su parte, las propiedades superficiales tienen cierta relevancia a la hora de fijar la referencia de un término de género natural, en tanto que las descripciones usadas en el bautismo inicial suelen involucrar descripciones que expresan esta clase de propiedades. No obstante, contrariamente a lo que afirma la Teoría Clásica, dichas descripciones no constituyen el significado del término de género natural y tampoco sirven para determinar su extensión en los usos posteriores del término, puesto que, como se ha visto, una entidad puede pertenecer al género aunque carezca de todas las propiedades superficiales relevantes. Lo fundamental en este sentido es que las instancias de un género natural compartan la misma propiedad estructural.

Una vez introducido el término (nombre propio o género natural) se inicia una *cadena histórica de comunicación* de la cual son eslabones cada uno de los usuarios de la expresión, quienes, por su parte, deberán tener la intención de usar la expresión con la misma referencia que tenía para aquellos de quienes lo recibieron. Puede observarse que

con el cambio de las descripciones propuestas por la Teoría Clásica a los encuentros causales encadenados hemos pasado de una explicación del vínculo referencial de carácter internista a una explicación externista. De este modo, si la Teoría Clásica exige que el hablante tenga conocimiento (interno) de esas descripciones para establecer dicho vínculo, ahora se propone una explicación de corte externista, que conecta la expresión con su *designatum* en función de cadenas históricas de comunicación externas a los usuarios del lenguaje.

2.2 *Las tesis de Putnam sobre la semántica de los términos de género natural y los experimentos mentales de la Tierra Gemela*

Analizadas las propuestas de Kripke, dedicaremos esta subsección al trabajo de Putnam de (1975c), de fundamental importancia para nuestra investigación. En este artículo, el autor pretende hacer explícitas las insuficiencias de la Teoría Tradicional del Significado de los términos de género natural, de corte internista y fregeano. La descripción que Putnam nos ofrece de esta teoría es la siguiente. Según la Teoría Tradicional, existen dos ingredientes básicos en el significado de nuestras expresiones, a saber, la *intensión* y la *extensión* o, empleando una terminología que habrá de resultarnos más familiar, el sentido y la referencia. La relación entre estos dos componentes semánticos se establece de tal modo que el sentido de un término, su intención, determina la extensión que le es propia en virtud de especificar una propiedad o una serie de propiedades que identifican dicha extensión. Putnam nos dice que, para Carnap, un partidario de esta Teoría Tradicional, la intención constituye un “criterio de pertenencia a la extensión (no sólo en el sentido de ‘condición necesaria y suficiente’, sino en el modo fuerte de modo de determinar si una cosa cae o no dentro de la extensión del término)” (1975c, p. 219). Es decir, que si dos términos compartieran intención, necesariamente habrían de designar el mismo conjunto de entidades. Putnam sostiene, además, que la Teoría Tradicional del Significado ha identificado la intención con un “concepto”. No obstante, hay que realizar algunas matizaciones. Entre los propios partidarios de esta Teoría Tradicional no existe consenso sobre el estatus ontológico de las intenciones o sentidos de nuestros términos lingüísticos. Por ejemplo, mientras que algunos autores como Locke mantienen que se trata de entidades

mentales,⁶¹ para Frege⁶² y Carnap⁶³ las intensiones tienen un carácter extramental e intersubjetivo. Sin embargo, en esta disputa existe un punto de común acuerdo: cualquiera que sea la “ubicación” de las intensiones, éstas han de ser elementos plenamente cognoscibles por la mente individual de aquellos hablantes competentes en el uso de los términos con tales intensiones. Repárese en que, en cualquier caso, la Teoría Tradicional del Significado preserva su tinte internista, ya que, por un lado, mantiene que el significado (i.e. la intensión) de un término de género natural es una entidad cognitivamente accesible a la mente individual y, por otro, afirma que la determinación de la referencia de los términos lingüísticos requiere la mediación de entidades internas así concebidas. En su artículo, Putnam recoge el contenido esencial de la Teoría Tradicional del Significado desglosándolo en la conjunción de dos tesis:

(TTSI) Conocer el significado de un término es estar en un cierto estado psicológico.

(TTSII) El significado de un término (entendido como su intensión) determina su extensión.

Hay que realizar una precisión acerca de la naturaleza de los estados psicológicos mencionados en la primera de las tesis de la Teoría Tradicional del Significado. Según Putnam, el estado psicológico en el que se encuentra un hablante que conoce el significado de un término es un estado acorde con el *solipsismo metodológico*. Éste es el nombre que recibe el enfoque teórico cartesiano que los valedores de la Teoría Tradicional del Significado han adoptado en sus análisis de los estados psicológicos. Dicho enfoque establece que sólo hemos de tomar en consideración aquellos estados psicológicos que no presupongan la existencia de un individuo distinto de aquél al que se atribuye tal estado. Para entender esto, pensemos, a modo de ejemplo, en dos estados

⁶¹ En (2008) p. 394, Locke afirma: “Resulta, pues, que el uso de las palabras consiste en que sean las señales sensibles de las ideas; y las ideas que se significan con las palabras, son su propia e inmediata significación.”

⁶² “Si alguien confunde los pensamientos con ideas, lo que él acepta como verdadero es, desde su propio punto de vista, el contenido de su conciencia, y ello no le concierne a nadie más en absoluto. Si él escuchara de mí la opinión de que un pensamiento no es una idea, ni siquiera podría discutirlo porque, de hecho, no le concierne. Así que la conclusión podría ser: los pensamientos no son ni cosas del mundo externo ni ideas. Debemos reconocer un tercer reino. Cualquier cosa que pertenezca a él tiene en común con las ideas que no puede ser percibido con los sentidos, pero tiene en común con las cosas que no necesita un propietario a cuyos contenidos de conciencia pertenezca” (Frege 1918, p.45).

⁶³ “El término “concepto” será usado aquí como una designación común para las propiedades, relaciones y entidades similares (incluyendo conceptos individuales [...] y funciones pero no proposiciones). Es especialmente importante resaltar, con respecto a este término, el hecho de que no debe ser entendido en un sentido mental, esto es, como si se refiriera a procesos de imaginación, pensamiento, concepción o semejantes, sino más bien a algo objetivo que encontramos en la naturaleza y que es expresado en un lenguaje por un designador de forma no sentencial” (Carnap 1975, p. 21).

psicológicos diferentes atribuidos a Juan: Juan siente dolor y Juan quiere a María. El primero de estos estados encaja perfectamente en las coordenadas del solipsismo metodológico, ya que la atribución a Juan de dicho estado no presupone ningún individuo distinto de Juan mismo. El segundo ejemplo, Juan quiere a María, es un caso diferente, pues parece que debemos presuponer la existencia de María, un individuo distinto de Juan, para dar cuenta de dicho estado. Por ello, si de acuerdo con el solipsismo metodológico quisiéramos examinar el estado psicológico en el que Juan se encuentra con respecto a María tendríamos que someter dicho estado a una reconstrucción y decir, por ejemplo, que el afecto que Juan dice sentir hacia María se dirige, en realidad, hacia sus propias representaciones mentales de María. Putnam propone una denominación para aquellos estados psicológicos que cumplen los requisitos del solipsismo metodológico, a saber, *estados psicológicos en sentido restringido* (o estrecho, como a veces se dice). Por el contrario, aquellos estados psicológicos que presuponen la existencia de un individuo distinto de aquél al que se atribuye dicho estado serían *estados psicológicos en sentido amplio*.

Estas observaciones ponen de manifiesto que el estado psicológico en el que se encuentra un hablante cuando conoce el significado de un término es, según la Teoría Tradicional del Significado, un estado psicológico en sentido restringido. Así, para caracterizar el estado psicológico de un sujeto cuando conoce el significado de un término “T” lo único que se requiere es conocer la intensión que este hablante asocia en su mente con dicho término. De esta manera, para establecer que dos hablantes se encuentran en el mismo estado psicológico, que especificamos como “conocer el significado de T”, debemos averiguar si estos hablantes asocian el mismo concepto o intensión con el mismo término. En caso afirmativo, ambos se encuentran en idénticos estados psicológicos. Además, si recordamos el segundo supuesto de la Teoría Tradicional del Significado que hemos recogido en (TTSII), a saber, que la intensión de las expresiones lingüísticas determina su extensión, la Teoría Tradicional del Significado está comprometida con un segundo hecho relativo a dos hablantes que asocian la misma intensión con el mismo término: sus estados psicológicos determinan exactamente la misma extensión para el término.

Putnam adopta una postura firme frente a la Teoría Tradicional del Significado en su artículo: “[a]rgüiré que no hay ninguna noción – no hablemos ya de la de significado – que satisfaga al mismo tiempo estos dos supuestos. El concepto tradicional de significado es un concepto que se basa en una teoría falsa” (1975c, p. 219). La estrategia

de Putnam para poner en evidencia los errores de la solución tradicional relativa al significado de los términos de género natural pasa por la elaboración de una serie de experimentos mentales que le ayudarán a mostrar que un mismo término, proferido por hablantes que se encuentran en el mismo estado psicológico en relación a dichos términos, podría tener distintas extensiones. Tengamos en cuenta, eso sí, que un argumento de esa clase sólo podría establecer que la conjunción de las dos tesis de la Teoría Tradicional del Significado es falsa, en la medida en que se sigue de dicha conjunción que dos estados psicológicos idénticos determinan la misma extensión para los términos relevantes, justo lo que Putnam intenta negar. Por tanto, si deseáramos sostener que sólo una de estas tesis es errónea y, en este sentido, responsable de la falsedad de esta consecuencia, tendríamos que basarnos en otras consideraciones independientes.

Dicho todo esto, conviene que atendamos al primero de los experimentos mentales de la Tierra Gemela que Putnam presenta para poner en cuestión la Teoría Tradicional del Significado. Este autor nos invita a imaginar que en algún lugar de nuestra galaxia existe un planeta, al que denominaremos “Tierra Gemela”, que es, salvo en cierto detalle que en seguida mencionaremos, enteramente idéntico al nuestro. Cada uno de los habitantes de la Tierra tiene en la Tierra Gemela una réplica por completo idéntica a él mismo, un duplicado que habla exactamente el mismo idioma que su doble en la Tierra, excepto en lo que concierne a un término lingüístico. Digamos ya, para deshacer el misterio, que aquel aspecto en el que difieren los entornos de la Tierra y de la Tierra Gemela consiste en que en este planeta lejano el líquido que corre por ríos y lagos, que calma la sed de los seres vivos, y que los hablantes de la Tierra Gemela denominan “agua” es un líquido con una composición química muy distinta de la del agua terrestre. No obstante, hay que aclarar que el líquido relevante de la Tierra Gemela comparte con el agua todas sus *propiedades macroscópicas*. En consecuencia, el líquido al que los habitantes de la Tierra Gemela denominan “agua” es, como el agua, un líquido inodoro, incoloro e insípido, a pesar de que las *propiedades subyacentes* o *microestructurales* del líquido en cuestión (su composición química) sean muy diferentes de las del líquido existente en la Tierra.⁶⁴ Podríamos hacer uso de la fórmula “XYZ” para abreviar la composición química del líquido llamado “agua” en la Tierra Gemela.

⁶⁴ Nótese que la metafísica de los géneros naturales de Kripke y de Putnam es muy similar, en el sentido de que el rol que el primero asigna a las propiedades superficiales equivale al que Putnam asigna a las

Dispuesto este escenario, Putnam desarrolla sobre él dos variaciones, aunque sólo nos ocuparemos en detalle de la segunda.⁶⁵ Este autor nos invita a retroceder en el tiempo hasta 1750. El sentido de fijar esta fecha radica en que por aquella época ni en la Tierra ni en la Tierra Gemela se conocía la composición de los líquidos que ambas comunidades lingüísticas denominan “agua”. Putnam nos pide imaginar a un hablante terrestre, al que llamaremos “Óscar₁”, y a su contrapartida en la Tierra Gemela, Óscar₂. Se sigue de la descripción de la Tierra Gemela como escenario-réplica de la Tierra (exceptuando el detalle que hemos mencionado) que Óscar₁ y su gemelo son réplicas exactas. Putnam dice al respecto que “se puede suponer incluso que Óscar₁ y Óscar₂ eran el uno el doble del otro en cuanto a su apariencia, sentimientos, pensamientos, monólogo interior, etc.” (1975c, p. 224). Óscar₁ y Óscar₂, en tanto que dobles exactos, compartirán necesariamente sus estados psicológicos en sentido restringido y, más concretamente, se encontrarán en el mismo estado psicológico cuando conozcan el significado del término “agua”, dado que ambos asociarán la misma intensidad con ese término. Es natural pensar que las propiedades recogidas en la intensidad relevante serán las propiedades macroscópicas del agua, puesto que en 1750 nadie conoce sus propiedades subyacentes. Si atendemos al criterio de identidad para la clase de estados psicológicos que consisten en conocer el significado de un término, ambos agentes se encuentran en el mismo estado, puesto que Óscar₁ y Óscar₂ asocian el mismo concepto-intensidad con el mismo término, pero, a pesar de ello, el término “agua”, proferido por hablantes que permanecen en *idénticos estados psicológicos*, *determina extensiones distintas* en un entorno y otro: Óscar₁ aplicará ese término al H₂O mientras que Óscar₂ lo aplicará al XYZ.

propiedades macroscópicas y que lo mismo ocurre con las propiedades estructurales y las propiedades subyacentes o, como a veces diremos, estructura subyacente.

⁶⁵ En la primera variación Putnam nos pide que imaginemos que una nave espacial terrestre llega a la Tierra Gemela. Es natural suponer que los habitantes de la Tierra que desembarquen en ese planeta considerarán, a primera vista, que el líquido inodoro, incoloro e insípido que llena los mares y ríos de la Tierra Gemela, es agua, dada su semejanza superficial con el líquido relevante de la Tierra. En esta medida, los tripulantes terrestres considerarían que el término “agua” tiene exactamente el mismo significado en ambos entornos. Ahora bien, supongamos que tras un estudio pormenorizado del líquido acuoso que los habitantes de la Tierra Gemela denominan “agua” los viajeros terrestres descubren que su composición química es por completo distinta de la del H₂O. Putnam nos dice que es plausible suponer que los tripulantes rectificarían su anterior juicio tras este hallazgo. En esta medida, dada la diferencia radical en la composición química de ambos líquidos, habrá serios inconvenientes en seguir considerando que “agua” significa lo mismo en la Tierra y en la Tierra Gemela. Tengamos en cuenta que al comienzo del artículo considerábamos que uno de los ingredientes de la noción del “significado” de un término quedaba constituido por su extensión. Por tanto, si dos términos (homófonos) designan distintos conjuntos de entidades parece verosímil afirmar que, al menos en cierto sentido, no significan lo mismo.

Repárese en cuál es la relación existente entre los dos supuestos de la Teoría Tradicional del Significado y la conclusión que Putnam nos invita a extraer a partir de su experimento mental. Si Óscar₁ y Óscar₂ son sujetos que permanecen en el mismo estado psicológico (que asocian el mismo concepto o intensión con el término “agua”) y, sin embargo, el término “agua” proferido por cada uno de ellos tiene, en cada caso, una extensión distinta, entonces la Teoría Tradicional del Significado no está en lo correcto al afirmar que los estados psicológicos constituyen la condición necesaria y suficiente para determinar la extensión de un término.

El segundo experimento mental que Putnam presenta toma pie, de nuevo, en el escenario de la Tierra Gemela.⁶⁶ Imaginemos ahora, dice este autor, que la diferencia existente entre nuestro planeta y la Tierra Gemela concierne a dos géneros naturales distintos: el aluminio y el molibdeno. En nuestro planeta, el aluminio, es decir, el material con el que están fabricadas, entre otros artefactos, las ollas y las sartenes, es un metal mucho más abundante que el molibdeno. Por otro lado, consideramos que estos materiales son superficialmente indistinguibles a simple vista. En cuanto a la Tierra Gemela, supongamos que allí las proporciones de ambos metales se hayan invertidas y, además, los objetos que en la Tierra se fabrican con aluminio están hechos de molibdeno en la Tierra Gemela. Supongamos, además, una segunda variación entre los entornos. En la comunidad lingüística de Óscar₂ los términos “aluminio” y “molibdeno” tienen sus extensiones permutadas. Esto supone que mediante el término “aluminio” los hablantes de la comunidad lingüística de Óscar₂ (él incluido) se refieren al molibdeno y, a la inversa, ellos emplean el término “molibdeno” para hablar del aluminio. Por último, imaginemos que Óscar₁ y Óscar₂ no son expertos en absoluto en el dominio de estos términos y que ninguno de ellos es capaz por sí mismo de diferenciar un metal del otro. Es de suponer que, en tanto que réplicas idénticas, cuando ambos hablantes se encuentran en el estado psicológico en que consiste conocer el significado de “aluminio”, están asociando el mismo complejo de propiedades en calidad de concepto o intensión con el mismo término, y que, en esta medida, sus estados psicológicos serán idénticos. No obstante, el término “aluminio” tiene una extensión distinta en función de qué agente lo profiera. Óscar₁, por su parte, hace uso de ese término para referirse al conjunto de los ejemplares de aluminio, mientras que su duplicado, Óscar₂, se refiere con él al conjunto de los ejemplares de molibdeno. El segundo experimento mental

⁶⁶ Putnam había presentado una primera versión de este experimento en su trabajo de (1970).

ofrece, de nuevo, una conclusión contraria a la que se sigue de la conjunción de los dos supuestos de la Teoría Tradicional del Significado, ya que muestra la posibilidad de que dos hablantes permanezcan en el mismo estado mental y que los términos relevantes involucrados en sus preferencias tengan extensiones distintas.⁶⁷ El “eslogan” mediante el cual Putnam recoge la intuición fundamental de los experimentos mentales planteados se ha hecho muy popular: “los significados no están en la cabeza” (1975c, p. 227). Dedicaremos el resto de la presente subsección a esclarecer qué quiere decir esta afirmación de Putnam, cuál es su relación con los dos supuestos fundamentales de la Teoría Tradicional del Significado y qué acercamiento alternativo propone este autor con respecto a las propiedades semánticas de los términos de género natural.

En primer lugar, es preceptivo hacernos cargo de la naturaleza de la conexión referencial existente entre nuestros términos de género natural y aquellas entidades extralingüísticas que constituyen sus extensiones. Si aceptamos las conclusiones que arrojan los experimentos mentales planteados, entonces parece que los conceptos individuales de Óscar₁ y Óscar₂ no son suficientes para determinar la extensión del término “agua”, como requería la Teoría Tradicional. El concepto que ambos asocian con el término “agua”, idéntico por hipótesis, recoge un patrón de propiedades macroscópicas que los dos líquidos comparten y que ambos hablantes conocen: carencia de olor, de color, de sabor, capacidad de saciar la sed, etc. Sin embargo, argumenta Putnam, el experimento pone en evidencia que los conceptos o intensiones así entendidos no pueden desempeñar el papel de instrumentos de determinación de la extensión de los términos de género natural que la Teoría Tradicional del Significado les había asignado. Putnam debe, pues, ofrecer una alternativa para responder a la cuestión de cómo se fija la extensión de dichos términos. Su propuesta toma pie en dos elementos: la *contribución del entorno natural* y la *contribución del entorno social*. En lo que respecta al primer factor, Putnam afirma que cuando deseamos fijar la extensión de un término en el idiolecto de un hablante podemos servirnos de dos procedimientos,

⁶⁷ A decir verdad, Putnam ofrece un tercer experimento mental que involucra los términos “olmo” y “haya”. Sin embargo, puesto que su estructura es muy similar a la del experimento diseñado en torno al uso de los términos “aluminio” y “molibdeno” creemos que es suficiente con dedicarle esta nota. El punto de partida del tercer experimento mental es la confesión de Putnam sobre su propia incapacidad para distinguir un olmo de una haya, lo cual le lleva a suponer que el concepto que asocia con ambos es idéntico (el concepto de árbol de hoja caduca, por ejemplo). Si construyéramos un entorno gemelo semejante al diseñado en el caso del aluminio y del molibdeno, podríamos pensar en un individuo enteramente idéntico a Putnam que, siguiendo los usos lingüísticos de su comunidad, llama “olmos” a las hayas y “hayas” a los olmos. No obstante, a pesar de que los conceptos que Putnam y su sosias asocian con ambos términos son enteramente idénticos, no estaríamos dispuestos a decir que cuando Putnam usa el término “olmo” esta expresión tiene el mismo significado que cuando lo usa su doble.

a saber, las *definiciones ostensivas* y las *definiciones operacionales*. Mediante las primeras, fijamos la extensión de un término mostrando al oyente un objeto o individuo que forme parte de dicha extensión. Por ejemplo, si nuestro interlocutor desconoce el significado del término de género natural “pirita” podemos emplear una porción de este mineral y mostrárselo, acompañando nuestro acto de la preferencia “esto es pirita”. Cabe emplear, asimismo, definiciones operacionales, por medio de las cuales fijamos en el idiolecto de un hablante la extensión un término de género natural “T” sirviéndonos de aquellas características que se asocian típicamente con su extensión.⁶⁸ El alcance de las definiciones ostensivas y operacionales, determinado por “la lógica de los términos de género natural” (Putnam 1975c, pp. 224-225), es tal que todo ejemplar de T así definido habrá de compartir las mismas propiedades subyacentes que los ejemplares paradigmáticos de T en nuestro entorno, ejemplares que, en el caso de la pirita, serán amarillos, brillantes y carentes de valor.⁶⁹ Tanto en las definiciones ostensivas como en las operacionales, la existencia de una naturaleza subyacente común en las entidades que pertenecen a la extensión de un término de género natural supone un factor irrenunciable a la hora de ofrecer una teoría correcta sobre la extensión de dichos términos.⁷⁰ Además, puede verse que en ambos procedimientos existe un componente indécico involucrado. En el caso de la definición ostensiva que nos propone Putnam, destaca la presencia del demostrativo “esto”, mientras que en la definición operacional tenemos la intención de fijar la extensión del término de género natural “T” apelando a la naturaleza subyacente compartida por todos los ejemplares que pertenecen a su extensión en *nuestro* entorno. Estas exigencias (identidad de naturaleza y componentes

⁶⁸ Esas características son, como veremos, los estereotipos y marcadores semánticos.

⁶⁹ Se ha señalado con frecuencia que hay una cierta inconsistencia entre el planteamiento del experimento mental de la Tierra Gemela y las descripciones que Putnam ofrece de las definiciones operacionales. Esa tensión existe en la medida en que la Tierra Gemela es un planeta lejano de nuestra galaxia (perteneciente al mundo actual) y, sin embargo, las definiciones operacionales que fijan la extensión de “agua” en el idiolecto de Óscar₂ tienen un alcance restringido a los ejemplares de agua *en su mundo* (en el mundo actual, que también es el mundo de Óscar₁). Aceptando esto, parece difícil concluir que el término “agua” tiene un significado distinto cuando es empleado por los habitantes de la Tierra Gemela. Para sortear este problema es frecuente tomar a la Tierra Gemela como un entorno contrafáctico. Nosotros emplearemos la expresión “nuestro entorno” en las definiciones operacionales con la esperanza de respetar la idea de Putnam de que en un entorno y otro la extensión del término “agua” difiere y, con ello, su pretensión de que estas definiciones fijan en los idiolectos de Óscar₁ y Óscar₂ extensiones distintas para esa expresión.

⁷⁰ Putnam indica que la presuposición de que los individuos que forman parte de la extensión de un término de género natural han de tener una naturaleza común compartida es una presuposición empírica. En tanto que empírica, tiene un carácter falible y, por tanto, la presuposición sería falsa si inadvertidamente el interlocutor de Óscar₁ está señalando un vaso de vodka mientras profiere la expresión “esto es agua”. La relación de identidad de naturaleza tiene un carácter teórico, dado que es competencia de la ciencia establecer en qué consiste exactamente la identidad de naturaleza de todos los miembros de la extensión de un término de género natural. Cf. Putnam (1975c), p. 225 y (1973b), pp. 122-123.

indéxicos) introducen al propio entorno en la fijación de la extensión de los términos de género natural y, además, son las responsables del carácter rígido de la designación del término “agua”.⁷¹ A este respecto, Putnam interpreta la definición ostensiva “*este líquido es agua*”, empleada para fijar la referencia del término “agua” en el idiolecto de Óscar₁, de modo que *para todo mundo posible* el término “agua” designa aquel líquido que posee la *misma naturaleza* que la porción de líquido designada en el mundo actual mediante el índice “este” que figura en la definición ostensiva.⁷² Putnam observa, en consecuencia, que la noción kripkeana de designación rígida puede extenderse a los términos de género natural como “agua” si adoptamos como criterio de identidad transmundo para lo designado por el término la naturaleza subyacente compartida por las porciones de agua de nuestro entorno (Putnam 1975c, p. 225).

Por lo que respecta a la contribución del entorno social, Putnam formula una hipótesis en (1975c), que denomina *hipótesis de la universalidad de la división del trabajo lingüístico*, según la cual en las comunidades lingüísticas es frecuente que los criterios para determinar qué individuos forman parte de la extensión de un término sólo sean conocidos por una minoría de hablantes, expertos en lo referido por la expresión. Parece intuitivo afirmar que un hablante competente en el uso del término “oro” puede no ser capaz por sí mismo de diferenciar una porción de este metal de un fragmento de pirita. Por ejemplo, este hablante sabe que el oro es un metal muy valioso, que en apariencia es amarillo y brillante, que su pureza se mide en quilates, etc., pero cuando quiere asegurarse de que un determinado objeto es (de) oro (i.e. pertenece a la extensión del término “oro”) confía en otros hablantes, concedores de los criterios para determinar qué es oro y qué no lo es. La hipótesis de la división del trabajo lingüístico es la forma en que Putnam da cuenta de un hecho fundamental sobre el uso del lenguaje,

⁷¹ Nótese que para defender el carácter rígido de la designación del término “agua” hemos de considerar, tal y como se ha recomendado en la nota 69, que la Tierra Gemela es un mundo posible y no un planeta lejano de nuestro mundo.

⁷² Ocurre algo similar con las definiciones operacionales que proporcionan una descripción de un término de género natural (“agua”, por ejemplo) formada por un marcador (“líquido”) y uno o varios estereotipos (“inodoro”, “inoloro”) que expresan propiedades distintivas normales (de carácter macroscópico) que típicamente poseen los ejemplares paradigmáticos de ese género natural. De acuerdo con Putnam (1975c, p. 232), el alcance de una definición operacional como la de “agua” es tal que necesariamente una entidad será una (instancia de) agua si es una porción de la misma sustancia cuyas instancias en nuestro mundo (actual) poseen las propiedades expresadas por los estereotipos. En cualquier caso, puede apreciarse que tanto las definiciones ostensivas como las definiciones operacionales han de interpretarse, según Putnam, de modo que el término de género natural así definido designe en *todo mundo posible* instancias que compartan *la misma naturaleza subyacente* que las instancias que en el mundo actual designamos haciendo uso del término. La diferencia entre ambos tipos de definiciones es que en el caso de las definiciones ostensivas nos servimos de una instancia del género relevante mientras que en las definiciones operacionales apelamos a propiedades características de ejemplares paradigmáticos de tales instancias, sin requerir la presencia de los mismos.

a saber, que todos los hablantes en mayor o menor medida presentan una *actitud deferencial* hacia otros miembros de su comunidad lingüística en el uso de muchos términos, esto es, en caso de duda acuden a otros hablantes mejor capacitados para determinar su aplicación.

Dicho esto, veremos de qué manera los resultados de los experimentos mentales diseñados por Putnam revelan que la contribución del entorno natural y social es un factor fundamental en la determinación de la extensión de nuestras expresiones. Cuando Óscar₁ y Óscar₂ fueron instruidos acerca del uso del término “agua” es probable que sus interlocutores se sirvieran de definiciones ostensivas y operacionales. El componente indéxico involucrado en ellas supone que la extensión del término “agua” en el idiolecto de Óscar₁ quedara fijada como aquel líquido que llenaba los vasos empleados para mostrarle a Óscar₁ que a *eso* se le denomina “agua”, un líquido que comparte la misma naturaleza subyacente con aquellas porciones del mismo que se consideran ejemplares paradigmáticos de agua en *su* entorno natural. Por tanto, las propiedades microestructurales de la sustancia relevante y la pretensión de que dichas propiedades deben ser compartidas por todos los ejemplares de agua, determina a qué se refiere Óscar₁ mediante el uso del término “agua”. Si, por hipótesis, la naturaleza subyacente del líquido que Óscar₂ y los hablantes de su comunidad llaman “agua” es radicalmente distinta de la naturaleza subyacente del agua (H₂O), entonces cuando Óscar₂ usa el término “agua” estará refiriéndose a un líquido distinto mediante el término en cuestión, a pesar de que los conceptos que asocia con él sean los mismos que asocia Óscar₁ con idéntico término.

En cuanto al experimento mental que involucra el uso de los términos “aluminio” y “molibdeno”, este escenario conjuga tanto la contribución del entorno natural del hablante como la contribución de su entorno social. Con respecto a esta última clase de contribución, este escenario incorpora un conjunto de expertos capaces de distinguir un metal del otro, al contrario que en el experimento anterior que, por desarrollarse en 1750, no incluía la existencia de agentes que pudieran saber que el agua y lo que en la Tierra Gemela se llama “agua” poseen distintas propiedades subyacentes y, por tanto, son sustancias distintas. Tanto Óscar₁ como Óscar₂ saben pocas cosas acerca del material que ambos denominan “aluminio” (quizá que las ollas y sartenes se fabrican con ese metal y que es más fácil de encontrar en sus respectivos entornos que el metal que denominan “molibdeno”). Desde luego, tanto uno como otro serían incapaces de distinguir una porción del metal que llaman “aluminio” de una porción de un material

idéntico a simple vista como es aquél que llaman “molibdeno”, aun cuando exista un grupo de hablantes en sus respectivas comunidades lingüísticas que sabe cómo hacerlo. La conclusión que nos invita a extraer Putnam de este segundo experimento consiste, recordemos, en mostrar que la intensión de un término de género natural, concebida al modo de la Teoría Tradicional del Significado, es insuficiente para determinar su extensión. Por ello, un mismo término, usado por dos sujetos que asocian con dicho término la misma intensión, podría tener una extensión distinta en función de qué sujeto lo use. ¿Qué determina, pues, la extensión de “aluminio” si no es la intensión que el hablante asocia con él? De nuevo, vemos que el término empleado por Óscar₁ tiene como extensión el conjunto de las porciones de aluminio, algunas de las cuales, quizá, fueron utilizadas como ejemplares en su aprendizaje del término, aunque sólo un subconjunto de los hablantes de su comunidad lingüística sepa determinar con precisión qué es aquello en virtud de lo cual un fragmento de metal pertenece a la extensión de aluminio. Con respecto a Óscar₂, nos resistimos a reconocer que la extensión del término “aluminio” tal como él lo usa sea la misma extensión que en su uso por parte de Óscar₁. Quienes le enseñaron el uso de ese término empleaban ejemplares de molibdeno para explicarle que eso era “aluminio”. Cuando complace a un hablante que se dirige a él con una petición como “acércame una olla de aluminio”, Óscar₂ le acercará una olla de molibdeno. A pesar de que Óscar₁ y Óscar₂ asocian las mismas propiedades a lo que uno y otro llaman “aluminio”, las porciones del material así denominado que han estado involucradas en sus encuentros tienen naturalezas esencialmente distintas, aunque sean otros hablantes quienes tienen la capacidad de conocer en rigor esa diferencia.

Una vez que hemos dado cuenta de las propuestas de Putnam acerca de la contribución del entorno natural y social como parte del mecanismo de designación de nuestros términos de género natural, es recomendable abordar sus sugerencias en lo relativo a la segunda dimensión semántica que estamos tomando en consideración: el significado en tanto que intensión o sentido. Sólo cuando hayamos expuesto en sus líneas fundamentales la alternativa de Putnam podremos explicar con propiedad su afirmación de que “los significados no están en la cabeza”. Para empezar, podemos recoger una observación que Putnam ofrece al hilo de su formulación de sus experimentos mentales. Estos experimentos demuestran que “identificar significado ‘en el sentido de intensión’ con *concepto* no puede ser correcto” (Putnam 1975c, p. 226). Si tomamos los conceptos como entidades epistémicamente accesibles por la mente individual y, además, consideramos que en el caso de los términos de género natural

estos conceptos recogen aquellas propiedades, generalmente de carácter macroestructural, que sirven para especificar a qué sustancias es correcto aplicar el término, los conceptos no pueden cumplir la función que la Teoría Tradicional del Significado les reserva. Óscar₁ y Óscar₂ tienen idénticos conceptos “en sus cabezas”, conocen idénticos significados – en el sentido de intensión – y, sin embargo, la extensión de los términos con los que asocian dichos conceptos difiere notablemente, lo cual nos llevaba a afirmar que el término “agua” no significa lo mismo para uno que para otro. Así las cosas, Putnam alega que los significados han de tener una constitución más compleja que la reflejada por la Teoría Tradicional del Significado y nos propone una nueva tesis al respecto conocida como “teoría vectorial del significado”. Para explicar esta teoría, tengamos en cuenta, en primer lugar, que un vector o tuplo es un conjunto ordenado de elementos, es decir, una serie de elementos en la cual los elementos que la componen así como la posición de cada uno de ellos determinan la identidad del vector. La idea de Putnam es que podemos ofrecer una definición del significado considerándolo un vector de, al menos, cuatro elementos, a saber, marcadores sintácticos, marcadores semánticos, estereotipos y extensión. Los *marcadores sintácticos* de un término son aquellas características sintácticas del mismo. Por ejemplo, entre los marcadores sintácticos del término “limón” estaría el rasgo de ser un término contable. Los *marcadores semánticos*, por su parte, son indicadores acerca de la categoría en la que podemos clasificar a las entidades que pertenecen a la extensión de un término, indicadores que consideramos poco susceptibles de revisión aunque su falsedad sea posible (Putnam 1975c, p. 267). Siguiendo con nuestro ejemplo, los marcadores semánticos de “limón” podrían ser “género natural” y “fruta”. Los *estereotipos* son aquellas características que un hablante medio asociaría típicamente con las entidades a las que se aplica el término y que consideramos que son rasgos menos centrales que los recogidos en los marcadores semánticos. En el caso de “limón” podríamos incluir, entre otros, “sabor ácido” y “color amarillo”. Por último, Putnam sostiene que en una descripción del significado de un término de género natural también debe figurar una *descripción de su extensión*, esto es, una descripción de aquella propiedad que recoge la naturaleza subyacente compartida por todas las entidades que forman parte de la extensión de un término. El ejemplo que ofrece este autor es “H₂O” como descripción de la extensión de “agua”; en el caso del término “limón” podemos suponer que describe una cierta secuencia de ADN. Es razonable que estas afirmaciones hagan pensar al lector que la descripción del significado de los términos que Putnam

propone es demasiado exigente. Así, las descripciones de la extensión de un término son altamente sofisticadas y presuponen un grado elevado de conocimiento por parte del hablante, amén de una ciencia madura que dé fundamento teórico a las composiciones químicas o los patrones genéticos. Ahora bien, Putnam nos dice que la competencia de un hablante con respecto al uso de un término de género natural va ligada a su conocimiento de los tres primeros componentes de la descripción de su significado. Empleando como ejemplo a los protagonistas del experimento mental, Óscar₁ y Óscar₂ son, cada uno de ellos, perfectamente competentes en sus usos del término “agua”, aunque desconozcan cuál es la composición química del líquido que designan mediante sus respectivos usos del término.

Examinadas las propuestas de Putnam acerca de la composición vectorial del significado y la participación del entorno natural y social en la determinación de la extensión de los términos de género natural, podemos hacernos cargo de la posición de este autor con respecto a los dos supuestos de la Teoría Tradicional del Significado. Por de pronto, recordemos que los resultados de los experimentos mentales pretendían mostrar la falsedad de una consecuencia extraída de ambos supuestos, a saber, que los estados psicológicos de los hablantes – entendidos como estados psicológicos en sentido restringido – determinan la extensión de sus términos. Esta tesis puede abreviarse diciendo que la teoría tradicional mantiene que “los significados están en la cabeza”, es decir, que dos sujetos individualmente idénticos y, por tanto, con “cabezas” idénticas, conocerían el mismo significado, puesto que dos estados mentales idénticos (consistentes en asociar una misma intensión con un mismo término) habrían de determinar una y la misma extensión para un término. Sin embargo, como vimos, “las cabezas” de Óscar₁ y de Óscar₂ son iguales, mientras que las extensiones y, por tanto, el significado de sus términos difieren. No obstante, aunque Putnam rechaza la conjunción de los dos supuestos de la Teoría Tradicional del Significado, no tiene por qué considerar que son ambos falsos (recordemos que para que una conjunción de supuestos sea falsa basta con que uno de ellos lo sea). De hecho, Putnam sostiene que el supuesto de que el significado-intensión determina la extensión es aceptable con ciertas reservas. La teoría vectorial del significado que propone este autor nos muestra que, en efecto, el significado de un término determina su extensión, aunque sólo en el sentido trivial de que la extensión de un término constituye uno de los ingredientes del significado del término.

Dicho todo esto, tenemos todos los elementos para comprender qué queremos decir cuando retratamos a Putnam como un valedor del externismo semántico. Recordemos, primeramente, que el carácter internista de la Teoría Tradicional del Significado descansaba en su afirmación de que aquello que determina la referencia o extensión de un término de género natural es un concepto o intensión, es decir, una entidad epistémicamente accesible al sujeto individual y perfectamente identificable sin apelar a nada distinto del sujeto mismo. Esto es, para establecer cuál es el significado del término “T₁” y averiguar si es diferente del significado del término “T₂” no habría que atender a nada que estuviera más allá de “los límites de la cabeza” del sujeto que conoce dicho significado. Si, además, el significado del término (cuya individuación es, para la teoría tradicional, función de factores internos al hablante) determina qué extensión le corresponde, los mecanismos de designación de nuestras expresiones serán, por ello, identificables a partir de elementos puramente individuales. Por lo que respecta a Putnam, hemos visto que su objetivo principal es mostrar que la intensión, identificada con un concepto e integrada en estados mentales en sentido restringido, no desempeña de manera apropiada el papel que la Teoría Tradicional del Significado le asigna, sino que es necesario tomar en consideración la contribución de factores externos al hablante, como la naturaleza físico-química de los géneros naturales de su entorno o los usos lingüísticos de otros hablantes de su comunidad lingüística, para capturar adecuadamente la extensión y el significado de los términos que ese agente usa. Si dos sujetos internamente idénticos están insertos en entornos distintos en el sentido relevante, la extensión y el significado del término “T” que ambos tienen en su vocabulario será diferente para cada uno de ellos, aun cuando no podamos señalar ninguna diferencia en los factores internos a dichos sujetos.

3. *Externismo psicológico (I): uso de los experimentos de la Tierra Gemela en Filosofía de la Mente*

Las propuestas de Putnam, que han quedado recogidas en lo fundamental en la sección precedente, dieron lugar a un acalorado debate en cuyo seno las tesis de *El significado de ‘significado’* se cuestionaron desde distintos puntos de vista. A pesar del interés que, indudablemente, tiene esta polémica, nosotros no vamos a examinarlas en profundidad

en este momento.⁷³ La razón es que nuestro planteamiento de las dos versiones del experimento mental de Putnam de la Tierra Gemela (centrada la primera en el término “agua” y la segunda en los términos “aluminio” y “molibdeno”), así como la explicación de sus propuestas en torno a la determinación externista de las propiedades semánticas, sólo constituye un medio para acceder al sentido de la polémica entre internistas y externistas en el ámbito de la psicología. Recordemos que la preocupación con la que iniciábamos este capítulo era la pregunta por las condiciones de identidad de nuestros contenidos mentales de carácter proposicional. Es decir, lo que deseamos saber es si los elementos internos al hablante bastan por sí mismos para determinar qué es lo que un hablante desea, cree, espera o teme o si, más bien, hay que atender a cuestiones adicionales relacionadas con la naturaleza de su entorno externo para individuar el

⁷³ Aunque a lo largo de los capítulos 3, 4 y, especialmente, en el 5 de este trabajo tendremos ocasión de discutir en profundidad algunos de los supuestos de este experimento mental, conviene indicar cuáles han sido los aspectos más polémicos del mismo. En este sentido, se ha puesto en duda la coherencia de las propias premisas de sus experimentos mentales, el compromiso de Putnam con una visión esencialista de los géneros naturales, su retrato de la “Teoría Tradicional del Significado”, el alcance de sus ataques a dicha teoría y, desde luego, la viabilidad de su propia alternativa. Por ejemplo, para algunos autores no es verosímil pensar que dos líquidos con propiedades microestructurales tan distintas como el agua y el líquido de la Tierra Gemela tengan idénticas propiedades macroestructurales, especialmente si pensamos que las primeras son causalmente responsables de las segundas (Segal 2000, p. 25). Otros autores consideran que defender la existencia de géneros naturales supone rehabilitar un modelo ontológico esencialista ya superado (Mellor 1977, p. 79). Se ha cuestionado, además, la afirmación de Putnam de que el agua es H₂O y sólo H₂O o, lo que es igual, su rechazo a admitir que el término “agua” tenga una extensión mixta y pueda referirse, así, al H₂O en la Tierra y al XYZ en la Tierra Gemela. Si ocurriera de esta manera, no existiría una diferencia en la extensión del término usado por Óscar₁ y Óscar₂ ni, consiguientemente, una diferencia en su significado del término “agua”. Autores como Searle niegan, por su parte, que las conclusiones de Putnam supongan un punto de ruptura fundamental con la teoría semántica tradicional. De este modo, sólo hace falta preguntarnos si acaso las definiciones ostensivas, mediante las cuales determinamos la extensión de los términos de género natural, podrían estar en un lugar distinto de la cabeza de los hablantes, en el sentido de que no tuvieran nada que ver con las intenciones (internas) de dichos hablantes, algo sumamente implausible. Además, la hipótesis de la división social del trabajo lingüístico es perfectamente asimilable por parte de una teoría semántica internista un tanto más sofisticada que la tradicional. Al fin y al cabo los estados mentales de los hablantes (de los hablantes expertos, eso sí) seguirían cumpliendo la función de determinar la extensión de los términos de género natural. Si un individuo ignorante desconoce con exactitud los criterios de aplicación de “T”, podríamos decir que domina “T” siempre y cuando presente una actitud deferente hacia los miembros relevantes de su comunidad lingüística. Y dicha actitud podría consistir perfectamente en un estado psicológico individual. En relación con esto, existe un reconocimiento por parte de autores de orientación internista de la existencia de “descripciones parásitas”, esto es, de descripciones que un individuo asocia con un término en las que se remite al uso de dicho término por parte de otros hablantes más competentes en este sentido. Consúltese al respecto el empleo que hace Searle en (1983), pp. 243, 245, 248, de la expresión “descripciones parásitas”. Strawson, en (1979), p. 182, afirma, por su parte, que “[...] las descripciones identificadoras, aunque no tienen que incluir una referencia a la propia referencia del hablante al particular en cuestión, pueden incluir una referencia a la referencia de otro a tal particular”. Otros autores internistas como Zemach consideran que el experimento configurado a partir de los términos “aluminio” y “molibdeno” no arroja en absoluto la conclusión de que Óscar₁ y Óscar₂ comparten un mismo estado psicológico aunque determinen distinta extensión para el término “aluminio”. Esto es así porque cada uno de ellos delega el conocimiento de los criterios de aplicación del término a conjuntos de hablantes situados en entornos distintos. Cf. Zemach (1976). Putnam acabará admitiendo esta conclusión en (Pessin/Golberg 1996, p. xix).

contenido de sus pensamientos. Mientras que en el ámbito del lenguaje la polémica internismo-externismo gira en torno a la determinación de las propiedades semánticas de nuestras expresiones, en el ámbito de la Filosofía de la Mente buscamos establecer qué factores determinan los contenidos de los pensamientos de los agentes, o, como a veces se dice, sus propiedades mentales. En esta tercera sección nos proponemos elucidar la relación que existe entre ambas cuestiones, para lo cual vamos a atender especialmente a las observaciones de aquellos autores que desde el externismo psicológico han encontrado en los argumentos de Putnam un apoyo para sus propias tesis.

Es natural, sin embargo, que surja una cierta perplejidad con respecto a la posibilidad de un aprovechamiento del experimento de Putnam a favor del externismo psicológico. Dicha perplejidad descansa en lo siguiente. Si los partidarios del externismo psicológico afirman que hay que tomar en consideración elementos del entorno externo para determinar de manera correcta la identidad de los pensamientos de los hablantes, estos autores tendrán un gran interés en mostrar que dos hablantes por completo idénticos en sus aspectos internos pueden tener contenidos psicológicos distintos, debido a que los entornos externos en los que cada uno de ellos está inserto son diferentes. No obstante, el punto de partida de los experimentos mentales de Putnam no parece ser, en principio, demasiado favorable para este tipo de pretensiones. Tengamos en cuenta que la efectividad del ataque de Putnam contra la conjunción de los dos supuestos de la Teoría Tradicional del Significado depende esencialmente de establecer la identidad de los estados mentales de Óscar₁ y Óscar₂ cuando conocen el significado de “agua”. Pero, si esto es así, ¿cómo podemos reutilizar los experimentos mentales de Putnam en provecho de una tesis que defiende que en ocasiones las diferencias en el entorno externo de los hablantes conllevan diferencias en sus estados mentales? Por lo que parece, si somos fieles a las propuestas de Putnam, las variaciones relevantes en el entorno externo de los hablantes conllevarían diferencias en la semántica de los términos que usan, pero no en sus estados mentales.⁷⁴

Además, existe una cierta interpretación del componente indécico involucrado en los términos de género natural que agudiza las dificultades que pueden tener los partidarios de emplear la argumentación de Putnam en favor del externismo

⁷⁴ Cf. Burge (1979), p. 108 y (1982), pp. 85 y ss.

psicológico.⁷⁵ Para verlo, tomemos en consideración las siguientes declaraciones de este autor:

“‘Agua’ es un término indéxico. ¿Qué quiero decir con esto? Si es un término indéxico, si lo que estoy diciendo es correcto, entonces ‘agua’ significa ‘lo que quiera que sea como el agua, que porte alguna relación de equivalencia, digamos la relación de líquido, con *nuestra* agua’. Donde ‘nuestra’ es, por supuesto, un término indéxico. Si esta es la forma en que la extensión de ‘agua’ es determinada, entonces el entorno determina la extensión de ‘agua’. Tanto si ‘nuestra’ agua es de hecho XYZ o H₂O” (Putnam 1974, p. 451).

Estas afirmaciones de Putnam podrían interpretarse del siguiente modo. El término “agua” lleva una regla semántica asociada, a saber, “‘agua’ significa lo que quiera que posea una naturaleza subyacente idéntica a la naturaleza de lo que nosotros denominamos ‘agua’”. Dicha regla, la misma para Óscar₁ y Óscar₂, nos da el significado de “agua” y determina para este término la extensión H₂O en la Tierra y la extensión XYZ en la Tierra Gemela. Leer así el comportamiento indéxico del término “agua” implicaría sostener que dicho término selecciona extensiones distintas en entornos distintos, al modo como un término indéxico (por ejemplo, “hoy”) selecciona extensiones distintas en contextos distintos en función de una regla semántica que permanece constante y que nos proporcionaría su significado. Pero al suponer que dicha regla está situada en “las cabezas” de Óscar₁ y Óscar₂, entonces los estados mentales de dichos sujetos permanecerían idénticos aun cuando la extensión de sus términos variara considerablemente. Por tanto, si consideramos que los pensamientos que involucran en sus contenidos conceptos expresados mediante términos de género natural tienen, ellos también, un comportamiento indéxico y, por tanto, pueden permanecer idénticos aunque remitan a distintos estados de cosas (la carencia de sabor del agua y la carencia de sabor del agua gemela, respectivamente), entonces las diferencias en el entorno externo de los agentes no implican una diferencia en la identidad de sus contenidos psicológicos. En consecuencia, el rédito que el externismo psicológico podría obtener de los argumentos de Putnam sería nulo.⁷⁶

Ahora bien, esta interpretación de la indexicalidad de los términos de género natural no es, según creemos, fiel a las propuestas de Putnam. Para demostrarlo, es preceptivo distinguir, por un lado, la presencia de un componente indéxico en las

⁷⁵ Cf. Salmon (2005a), pp. 101-104.

⁷⁶ Cf. Burge (1982), p. 87.

definiciones que fijan la extensión de estos términos y, por otro, el comportamiento semántico indéxico *tout court* de los términos de género natural. Mientras que Putnam acepta lo primero, no pensamos que sus tesis respalden lo segundo. Si el comportamiento semántico de los términos de género natural fuera asimilable al comportamiento de los términos indéxicos, entonces un término de género natural tendría un significado constante (la regla semántica que enunciamos), pero determinaría extensiones distintas en distintos entornos. Sin embargo, en (1975c) Putnam no considera que el significado de “agua” sea el mismo cuando el término es usado por Óscar₁ que cuando es usado por Óscar₂. De hecho, la fuerza de sus conclusiones al plantear el escenario de la Tierra Gemela descansa en la intuitiva reticencia a admitir que “agua” significa allí lo mismo que en la Tierra, a pesar de que las definiciones operacionales y ostensivas que determinan la extensión de ese término en un entorno y otro sean expresadas mediante preferencias homófonas. Es cierto que, en tanto que la extensión constituye uno de los elementos del vector significado de los términos de género natural como “agua”, la regla que determina la extensión nos proporciona una parte de lo que “agua” significa para uno y otro hablante. Pero esto no implica que esta regla semántica constituya, por sí misma, el significado del término ni, desde luego, que la regla semántica nos proporcione una definición del término. Por tanto, el término “agua” no sólo tiene extensiones distintas en entornos distintos, sino que la preferencia de Óscar₁ cuando dice “el agua es insípida” no significa lo mismo que una preferencia homófona de Óscar₂.

El próximo paso que debería dar un valedor del externismo psicológico que quisiera utilizar el experimento mental de Putnam consiste en demostrar que, en la medida en que uno y otro hablante no expresan lo mismo mediante estas preferencias homófonas, tampoco piensan lo mismo. Lo único que parece impedir esa afirmación es la premisa del experimento mental de Putnam que estipula que los gemelos permanecen en idénticos estados psicológicos cuando entienden el término “agua” tal como se usa en sus respectivos planetas. Sin embargo, recordemos que la identidad de los estados psicológicos de ambos agentes tiene como presupuesto el marco del solipsismo metodológico, esto es, el programa psicológico que restringe su análisis a los estados psicológicos que no presuponen ningún individuo distinto de aquél al que se le atribuyen dichos estados. A juzgar por las declaraciones de Putnam, parece que este autor sólo asume el solipsismo metodológico en tanto que éste constituye la pauta de análisis presupuesta por la teoría semántica tradicional que es objeto de crítica en sus

trabajos y, de hecho, tras su descripción de dicha pauta, Putnam se expresa del siguiente modo: “tres siglos de fracaso de la psicología mentalista es una tremenda evidencia en contra de esta forma de proceder” (1975c, p. 221). Esto pone de manifiesto que Putnam asume provisionalmente este programa para criticar cierta tendencia en la teoría tradicional del lenguaje, a saber, la tendencia a centrarse en el sujeto, considerado de forma aislada o puramente individual (1975c, p. 271).

A partir de esto, el defensor del externismo psicológico podrá proponernos un acercamiento al escenario del experimento mental de (1975c) que se deshaga de las restricciones impuestas por el solipsismo metodológico, el cual, como vemos, no es aceptado por Putnam. Para ver de qué modo podría realizarse este acercamiento, tomemos en consideración, de nuevo, el escenario de la Tierra Gemela. Por comodidad, podemos suponer que se trata de un entorno contrafáctico, esto es, hablaremos de la Tierra Gemela como un mundo posible enteramente idéntico al nuestro salvo por el aspecto que ya conocemos: el líquido acuoso macroscópicamente indistinguible del agua que corre por sus ríos y lagos es XYZ y no H₂O. De nuevo, pensemos en un hablante del mundo actual, Óscar₁, y en su réplica contrafáctica, Óscar₂, del cual diremos que es enteramente idéntico a Óscar₁ en todos sus aspectos intrínsecos. Supongamos que Óscar₁, el hablante del mundo actual, profiere la oración “el agua es húmeda”. Atendiendo a nuestras prácticas ordinarias de atribución de contenidos mentales, es natural reconocer en la mente de Óscar₁ la presencia de una creencia a ese respecto, dado lo cual diremos que Óscar₁ cree que el agua es húmeda. Ahora supongamos que Óscar₂, hablante del castellano gemelo, emite una oración homófona: “el agua es húmeda”. ¿Podemos atribuirle la misma creencia que a Óscar₁? Una serie de consideraciones desaconsejan una respuesta afirmativa a esta cuestión. Desde el momento en que aceptamos, con Putnam, que el término “agua” es un designador rígido, las oraciones proferidas por Óscar₁ y Óscar₂ que involucran ese término tendrán un significado distinto, ya que la extensión del término “agua” en castellano será el H₂O (en todos los mundos posibles) y la extensión del término “agua” en castellano gemelo será XYZ (en todos los mundos posibles). A consecuencia de esto, si atendemos a la preferencia por parte de Óscar₁ de la oración “el agua es húmeda”, esa oración será verdadera si y sólo si el agua es húmeda. En cambio, si aceptamos que el líquido designado mediante el término “agua” en la Tierra Gemela *no es agua*, cuando Óscar₂ dice “el agua es húmeda” se está refiriendo a un líquido distinto y, por tanto, esa oración, proferida por él, tendrá el valor de verdad verdadero sólo si ese líquido distinto (también

denominado “agua” por los habitantes de su entorno) tiene la propiedad de ser húmedo. Vemos, pues, que las condiciones de verdad de las oraciones proferidas por uno y otro agente son distintas. Es más, asumiendo que el líquido acuoso gemelo no es agua, parece natural decir que Óscar₂ nunca ha visto una porción de agua, jamás ha bebido ese líquido, nunca se ha bañado en agua y ni siquiera ha oído hablar de ella. Si esto es así, ¿cómo podría tener creencias relativas a la humedad del agua? Resultaría extremadamente contraintuitivo, por tanto, decir que las creencias de Óscar₁ y de Óscar₂ tienen el mismo contenido.⁷⁷

En lo que resta de este primer capítulo denominaremos a esta remodelación del experimento mental a favor del externismo psicológico “experimento^{ep} de la Tierra Gemela”, para diferenciarla de la formulación original, aunque a partir del segundo capítulo hablaremos, simplemente, de “experimento de la Tierra Gemela” para referirnos a esta reconstrucción. Sobre dicha remodelación han de hacerse dos observaciones. En primer lugar, esta reconstrucción apoya la tesis de los defensores de la individuación externista de los contenidos mentales, quienes sostienen que la identificación de dichos contenidos ha de atender a factores externos a los propios hablantes. Si hemos estipulado que Óscar₁ y Óscar₂ son intrínsecamente idénticos y, sin embargo, concluimos que los contenidos de sus pensamientos difieren (el primero cree que el agua es húmeda y el segundo no), entonces las condiciones de identidad de sus respectivos contenidos han de residir, al menos en parte, en su entorno externo. Dicho de otra manera, parece que la diferencia relevante entre sus entornos, la composición química del líquido que ambos llaman “agua” es la única responsable de que nos veamos obligados a establecer que uno y otro tienen creencias distintas, puesto que, por hipótesis, Óscar₁ y Óscar₂ son internamente idénticos.⁷⁸ La segunda observación que es interesante realizar sobre esta reconstrucción es que el propio Putnam terminó por dar la razón a estos autores que extrapolaron sus conclusiones semánticas externistas al ámbito de la psicología, diciendo así que “[...] he llegado a pensar que esto [la afirmación de que los estados psicológicos de Óscar₁ y Óscar₂ son idénticos] concede demasiado a la

⁷⁷ De una manera muy expresiva McDowell afirma que la consecuencia que Putnam extrajo de sus experimentos mentales, recogida en su famosa frase “los significados no están en la cabeza”, debería haber sido, más bien, que “la mente no está en la cabeza”, es decir, que la identidad de un estado mental es algo que en ocasiones sólo puede establecerse a partir de lo que está “fuera” de la cabeza de los individuos. Cf. McDowell (1992), p. 306.

⁷⁸ Una reconstrucción del experimento mental de Putnam concerniente a los términos “aluminio” y “molibdeno” desde el punto de vista del externismo psicológico puede consultarse en Burge (1979), p. 167 y (1989b), p. 316.

idea de que la mente puede ser considerada como un teatro privado (situado dentro de la cabeza)” (Pessin/Goldberg 1996, p. xviii).⁷⁹

A modo de recapitulación y con el fin de arrojar mayor claridad sobre la cuestión que nos ocupa, vamos a resumir la pauta de desarrollo de los experimentos mentales de la Tierra Gemela, tanto en la versión semántica que Putnam presenta en (1975c), como en su versión psicológica. En ambos casos, estipulamos un entorno, ya se trate de un planeta lejano o de un mundo posible, que es enteramente idéntico al nuestro salvo por un aspecto: la diferente composición química del líquido que en ambos entornos se denomina “agua”. Es de suma importancia que esa diferencia sea desconocida para los protagonistas del experimento mental; por ejemplo, Óscar₁ y Óscar₂ ignoran la composición química de lo que ambos llaman “agua”.⁸⁰ Así, el desconocimiento de la diferencia relevante entre los entornos nos da una de las claves para establecer con mayor precisión en qué consiste el diferencial interno-externo situado en el centro de la polémica entre los internistas y los externistas. En la medida en que ignoran la diferencia en sus entornos, Óscar₁ y Óscar₂ residen en entornos enteramente similares *por lo que a ellos respecta* (cf. Farkas 2003). Es frecuente expresar esta idea afirmando que Óscar₁ y Óscar₂ viven en *mundos fenomenológicamente idénticos* o *subjetivamente equivalentes* (Ebbs 2005, p. 239). Los factores internos cuya identidad estipulamos en ambos hablantes son aquellos componentes mentales que uno y otro comparten, como la cualidad o apariencia de sus experiencias subjetivas.⁸¹

⁷⁹ De hecho, en (1992) Putnam afirma que las conclusiones obtenidas gracias al experimento mental de la Tierra Gemela pueden ser usadas para mostrar que la comprensión funcionalista de los estados mentales (en tanto que individuados por su rol causal, de carácter interno al individuo) no es apropiada. En este trabajo Putnam considera que los argumentos desarrollados en (1975c) refuerzan la idea de que “el contenido de nuestras creencias y deseos no está determinado por propiedades individualistas del hablante, sino que depende de hechos sociales sobre la comunidad lingüística, de la cual es un miembro el hablante, y de hechos sobre el entorno físico del hablante y de la comunidad lingüística” (Putnam 1992, p. 443). Como puede apreciarse, aunque el planteamiento original del experimento de la Tierra Gemela en (1975c) no permitía reforzar, de suyo, una visión externista de los estados mentales de sus protagonistas, debido a que se estipulaba una identidad de partida entre dichos estados al estar estos considerados en el marco del solipsismo metodológico, las observaciones de Putnam en trabajos posteriores ponen de manifiesto que este autor acepta y se adhiere a la pretensión de emplear dicho experimento (con las modificaciones convenientes) para apoyar una concepción externista en el ámbito psicológico.

⁸⁰ Crane afirma que ésta es la pauta fundamental para diseñar cualquier experimento mental a favor del externismo. El internista tendría que demostrar que “no podría ser el caso de que todo les pareciera idéntico a los sujetos gemelos aunque las cosas a las que se refieren cada uno de sus pensamientos fueran diferentes” (Crane 2001, p. 123).

⁸¹ Al hacer esta consideración nos estamos distanciando del empleo por parte de algunos autores del experimento mental de la Tierra Gemela para mostrar la inviabilidad de una cierta versión de la teoría de la identidad mente-cerebro. Si consideramos que Óscar₁ y Óscar₂ son, por hipótesis, individuos con estados neuronales idénticos y, sin embargo, sus pensamientos difieren, queda deslegitimada una identificación reduccionista entre estados neuronales y estados mentales. No obstante, pensamos que la

El externista en psicología hace hincapié en el hecho de que las respectivas actitudes psicológicas de Óscar₁ y Óscar₂ se dirigen hacia contenidos esencialmente distintos. Esta diferencia objetiva queda incorporada en la historia de los encuentros en los que uno y otro han estado involucrados en sus distintos entornos. Si el líquido de la Tierra Gemela no es agua, Óscar₂ jamás ha visto, tocado o saboreado el agua y, por tanto, parece extremadamente improbable que esté atribuyendo humedad a ese líquido cuando dice “el agua es húmeda”. De aceptar esto, tendremos que concluir, como pide el externista, que al menos en ocasiones parece necesario atender a elementos externos, de carácter extra-mental, para ofrecer una explicación apropiada de la identidad de los contenidos mentales de los hablantes.

4. *Externismo psicológico (II): el anti-individualismo de Burge.*

En esta sección nos haremos cargo de una serie de experimentos mentales planteados por Tyler Burge con el objetivo de aportar sustento teórico al externismo psicológico. En su artículo de (1979) Burge incide en la relevancia de los factores sociales en el estudio del lenguaje y de la mente de los hablantes, factores que, en su opinión, han sido olvidados o relegados a un segundo plano en los acercamientos tradicionales a estas cuestiones. Frente a esa clase de enfoques, que este autor califica de “individualistas”, nos propone un programa distinto, que denomina *anti-individualismo* o, como a veces se dice, *externismo social*. Por otro lado, a pesar de ser uno de los autores que destacó el carácter limitado de las conclusiones de Putnam en (1975c), Burge reconoce el parentesco existente entre sus propuestas y las de este autor⁸² y propone la elaboración de una serie de experimentos mentales en la línea de Putnam a favor de la individuación externista de los contenidos mentales. En concreto, Burge quiere mostrar que el contexto social de los hablantes ha de ser tomado en consideración para dar una explicación apropiada de nuestros hábitos de atribución de creencias, defendiendo la idea de que la identidad de los contenidos de un hablante depende esencialmente de los usos lingüísticos que ha establecido la comunidad de hablantes en la que él está inserto.

Con el fin de entender esta tesis adecuadamente, conviene atender a los experimentos mentales que Burge ofrece en (1979), prestando especial atención al

identidad neuronal de los gemelos es colateral al argumento de Putnam y que sus intenciones originales siguen más bien nuestro planteamiento.

⁸² Cf. Burge (1979), p. 107; (1993a), p. 300 y (1993b), p. 350.

primero de ellos, centrado en el uso del término “artritis”, ya que es el experimento que mayor éxito histórico ha tenido. En este experimento Burge nos invita a imaginar a un hablante del castellano, a quien, siguiendo la tradición, llamaremos “Bert”, que tiene una serie de creencias en torno a la artritis. Bert cree que padece artritis desde hace varios años, que es una enfermedad muy dolorosa aunque preferible al cáncer, que el endurecimiento de las articulaciones es un síntoma propio de la artritis y que existen varios tipos de artritis. Este hablante cree, además, que sus padecimientos artríticos son responsables del dolor que siente en el muslo. Sin embargo, está en un error, puesto que la artritis es una enfermedad reumatoide que afecta únicamente a las articulaciones y en ningún caso provoca el tipo de dolor muscular que Bert siente. Supongamos, asimismo, que este hablante acude al médico para transmitirle su temor de estar padeciendo artritis en el muslo. No obstante, el doctor tranquiliza a Bert explicándole que la artritis es una patología que únicamente afecta a las articulaciones. Bert acepta la corrección del doctor y pregunta entonces por la causa de su dolor muscular.

Una vez planteada esta parte del experimento, podemos hacernos una pregunta: ¿es legítimo atribuir a Bert creencias sobre la artritis? En principio, no parece existir ningún inconveniente. Este hablante no es un completo ignorante respecto al uso del término “artritis”. De hecho, tiene una serie de creencias verdaderas sobre esta enfermedad y acude al experto relevante para resolver sus dudas acerca de ella. Parece, por tanto, que si reconocemos a Bert la capacidad de pensar sobre la artritis aun cuando algunos de sus pensamientos sean falsos, no hay razón para negarle la posesión de un concepto correspondiente. La conclusión más importante que podemos extraer de la conversación entre Bert y su médico en torno a la artritis es que ambos hablantes comparten el mismo concepto. Desde luego, nadie pondría en duda la atribución del concepto de artritis al doctor y, por lo que vemos, parece razonable pensar que también Bert lo posee. Todo ello muestra que un hablante puede comprender de manera incompleta una noción sin que esto impida que podamos atribuirle actitudes mentales acerca de la misma.

El fulcro de la argumentación de Burge es que cualquier otra interpretación alternativa de la situación cognitiva de Bert resulta mucho más forzada. Así, supongamos que alguien niega que podamos atribuirle a Bert la noción de artritis. Según esto, el concepto que el doctor posee (que es indudablemente el concepto de artritis) y el concepto que posee Bert son nociones distintas. Afirmar tal cosa tendría como consecuencia que un uso inadecuado de un término lingüístico – o una

comprensión incompleta de ese uso por parte de un hablante – revelaría que ese sujeto posee un concepto idiosincrásico, correspondiente al término lingüístico que sólo domina en parte. Además, si aceptamos esta interpretación, la descripción de la conversación entre Bert y el médico sería muy diferente. En vez de sostener que ambos hablantes comparten el mismo concepto de artritis y que Bert, gracias a su médico, ha corregido el uso que hacía del mismo, diríamos, más bien, que antes de que dicha conversación tuviera lugar ambos estaban en lo cierto con respecto a sus peculiares conceptos. Esto es, Bert aplicaba correctamente el término “artritis”, teniendo en cuenta que este término expresaba para él un concepto distinto. Sin embargo, Burge afirmaría que nuestras intuiciones nos impiden seguir por este camino. Por un lado, si tuviéramos que adscribir un concepto peculiar de X a cada uno de los hablantes que entienden parcialmente el uso de una expresión “X”, la comunicación entre los hablantes se complicaría hasta extremos insospechados, dada la presencia de innumerables conceptos idiosincrásicos en sus mentes. Por otro lado, el carácter normativo del lenguaje y la existencia de especialistas en el conocimiento de esas normas dentro de las comunidades lingüísticas vuelven razonable la actitud de Bert, quien, en el experimento mental de Burge, admite la corrección del doctor y deja de aplicar el término “artritis” a su dolor de muslo. Ante esta reacción parece más natural decir, con Burge, que antes de la conversación con el médico nuestro paciente aplicaba mal el concepto de artritis, más bien que decir que este sujeto aplicaba perfectamente su peculiar concepto, aunque su reconocimiento de la autoridad del médico le lleva a adquirir uno nuevo.

Visto esto, damos paso a la segunda parte del experimento mental. Burge nos pide imaginar una situación contrafáctica cuya única variación respecto de la situación actual consiste en que en la comunidad lingüística de Bert el término “artritis” se aplica no sólo a la dolencia que denominamos “artritis”, sino también a ciertos padecimientos reumatoides que afectan a los músculos. En esta situación, Bert tendría exactamente la misma constitución “interna” que en el mundo actual. Esto es, sus estados físicos serían idénticos, su historia fisiológica también, sus disposiciones a asentir ante preguntas como “¿padeces artritis?”, “¿crees que la artritis es una enfermedad dolorosa pero menos penosa que el cáncer?”, “¿piensas que tienes artritis en el muslo?” serían igualmente favorables. Bert, en esta situación, acude al médico y le expresa sus temores mediante la siguiente preferencia: “doctor, creo que sufro de artritis en el muslo”. En este caso, el médico da crédito al paciente y le receta algunos medicamentos para aliviar su dolor. El motivo es que en la comunidad lingüística contrafáctica se usa el término

“artritis” para designar un trastorno fisiológico distinto de la artritis y, en consecuencia, los síntomas que Bert atribuye a la enfermedad que denomina de ese modo encajan por completo en el cuadro clínico de la dolencia establecido en la comunidad contrafáctica.

Las cuestiones que nos interesa plantear en esta situación son dos. En primer lugar, nos preguntamos si los conceptos que Bert y su doctor expresan mediante su uso del término “artritis” en esta situación contrafáctica son o no los mismos. Desde luego, es natural responder de manera afirmativa. Incluso aquellos que mantenían sus reservas para con la situación actual no tienen motivos para ofrecer la misma respuesta en este caso: el uso que hace Bert del término “artritis” se adecua por completo al que hace su médico. La segunda pregunta es mucho más importante. Queremos saber qué concepto expresa Bert en la situación contrafáctica cuando usa el término “artritis”. Podemos plantear esta cuestión de manera indirecta: Bert, en la situación contrafáctica, ¿cree que tiene artritis? Si consideramos que la artritis es una enfermedad que sólo afecta a las articulaciones, entonces parece difícil admitir que Bert cree que tiene artritis en la situación contrafáctica, a pesar de que él siga empleando el término “artritis” para hablar de su padecimiento. Dicho de otra manera, si la enfermedad que en nuestra comunidad lingüística llamamos “artritis” es una enfermedad distinta de aquella que los hablantes de la comunidad contrafáctica denominan haciendo uso del mismo término, entonces no parece correcto mantener que los hablantes de la comunidad contrafáctica (incluido el Bert contrafáctico) creen que tienen artritis cuando dicen de alguien “que tiene artritis”.

La tercera y última parte del experimento mental nos permite extraer conclusiones favorables al anti-individualismo. En ella, Burge dirige su atención a Bert. El autor había estipulado que este hablante permanece intrínsecamente idéntico en ambas situaciones: su fisiología y su historia interna no varía entre un escenario y otro y sus disposiciones a asentir a las mismas oraciones son idénticas. Bert, por tanto, permanece idéntico *si lo consideramos individualmente*. Ahora, lo interesante es reparar en el hecho de que en la situación actual parecía plausible atribuirle el concepto de artritis mientras que negamos que posea dicha noción en la situación contrafáctica. Están dadas, pues, las premisas para obtener una conclusión anti-individualista, ya que la diferencia que detectamos en las creencias de Bert entre un escenario y otro no puede provenir de diferencias en las propiedades internas de dicho hablante, al haber estipulado su identidad en este sentido. Son los diferentes usos lingüísticos convencionalmente establecidos para el término “artritis” lo que determina la diferencia entre las creencias

de Bert en las dos situaciones consideradas. Estos factores de carácter social, externos al sujeto son, por tanto, los elementos responsables de que Bert reciba distintas atribuciones psicológicas en esos dos casos.⁸³

Nótese que al igual que en el experimento^{ep} de la Tierra Gemela, el argumento de Burge saca partido de un cierto componente de ignorancia por parte de los protagonistas de sus experimentos. Bert desconoce en la situación actual que la artritis es una enfermedad que sólo afecta a las articulaciones y, por tanto, alberga la creencia (errónea) de que su dolor de muslo está provocado por esa enfermedad. Si dominara por completo el significado del término “artritis”, como presuponemos que lo domina su médico, nunca le habría comunicado a éste su pensamiento de que tiene artritis en el muslo. En la situación contrafáctica, Bert aplica correctamente el término “artritis” atendiendo al uso de esta expresión legitimado por los expertos de su comunidad lingüística, entre los que se encuentra su doctor que, en este caso, confirma los temores de Bert. En cualquiera de las dos situaciones, la enfermedad que denomina “artritis” es, para Bert, la misma enfermedad *por lo que a él respecta*. Por ello, también podríamos decir que, al igual que Óscar₁ y Óscar₂ viven en mundos subjetivamente indiscernibles, la experiencia interna de Bert, su fenomenología, permanece inalterada en la situación contrafáctica. Lo que Burge y los valedores del experimento^{ep} de la Tierra Gemela quieren demostrar es que, a menudo, las adscripciones de contenidos mentales a los hablantes han de contar con factores que trascienden su fenomenología. Por decirlo de otra manera, los experimentos mentales de este tipo muestran que las condiciones de

⁸³ Burge desarrolla en (1979) otros experimentos mentales acerca de los conceptos expresados por otros términos generales como “pecho de res”, “contrato”, “rojo”. Su estructura es exactamente la misma que en el caso del concepto de artritis. Se parte de un individuo del mundo actual, hablante del castellano, que tiene pensamientos con contenidos que involucran los conceptos expresados por estos términos aunque algunos de estos pensamientos sean erróneos. Después, Burge describe la conducta de estos sujetos de manera que resulte plausible atribuir al individuo la posesión de los conceptos relevantes. A partir de esta situación, se diseña una variación contrafáctica de cada uno de estos escenarios en las cuales las propiedades internas de los individuos permanecen constantes, mientras que el significado que se establece socialmente para el término en cuestión es diferente. La clave reside en que aquello que los hablantes del escenario contrafáctico denominan “pecho de res” no es pecho de res, que el término “contrato” no se aplica a los contratos y que el color que llaman “rojo” no es el rojo. De ser así, el contenido de los pensamientos del protagonista del experimento cambia en la situación contrafáctica, dado que no podremos atribuirle contenidos sobre el pecho de res, los contratos o el color rojo cuando usa dichos términos, contrariamente a lo que ocurría en la situación actual. Sin embargo, al haber estipulado la identidad interna del protagonista del experimento en ambas situaciones, entonces la diferencia en sus pensamientos sólo puede deberse a la variación del significado socialmente establecido de sus palabras. Desde luego, para que esta variación deje inalteradas las propiedades intrínsecas del individuo es un requisito fundamental que el individuo desconozca en ambas situaciones el significado *completo* del término relevante.

identidad de los contenidos que atribuimos pueden estar fuera del alcance epistémico de los sujetos que ameritan estas atribuciones.⁸⁴

A pesar de las semejanzas existentes entre el planteamiento de Burge y el planteamiento de Putnam rediseñado a favor del externismo psicológico, podemos percibir algunas diferencias (cf. Donnellan 2003). Para empezar, las tesis de Burge tienen un alcance mucho mayor. Como vimos, Putnam centra su estudio en cierto tipo de términos generales, a saber, los términos de género natural, aunque también extiende sus consideraciones – aunque con menor plausibilidad – a los términos de género artificial y términos socio-legales. Sin embargo, Burge tiene en mente un proyecto más ambicioso. La determinación externista puede afectar a cualquier pensamiento cuyo contenido involucre conceptos aplicables a clases públicas de objetos, propiedades o eventos,⁸⁵ lo cual supone en la práctica que todos los términos generales, cuyo uso es susceptible de recibir una variación contrafáctica del tipo reseñado, expresan conceptos cuyas condiciones de identidad dependen parcialmente del entorno social del sujeto, entorno que, por la vía de los expertos relevantes, se encargará de establecer la aplicación correcta de los términos. En relación con esto, encontramos una segunda diferencia entre las conclusiones de Burge y las que obtenemos al formular el experimento^{ep} de la Tierra Gemela. Mientras que para Putnam los expertos mencionados en la hipótesis de la división social del trabajo lingüístico son fundamentalmente los científicos que conocen los criterios de aplicación de un término de género natural a las entidades que pertenecen a su extensión, los expertos, para Burge, forman parte de colectivos mucho más variados. Así, en la medida en que la determinación externista afecta a la práctica totalidad de los términos generales, para Burge serán expertos en el uso de una expresión aquellos hablantes que dominan las

⁸⁴ Sería interesante preguntarnos hasta qué punto un individuo puede desconocer el uso del término “T” sin que ello suponga un impedimento para atribuirle el concepto de T. Burge considera que las atribuciones de contenido son perfectamente falibles y, por tanto, si un hablante poseyera el término “T” en su vocabulario y decidiéramos atribuirle un concepto correspondiente, pero, tras someterlo a un cuestionamiento posterior, viésemos que sólo asocia creencias erróneas con él, sería perfectamente legítimo rectificar nuestro anterior juicio. Burge afirma que el mínimo imprescindible para atribuir un concepto a un individuo consiste en su capacidad de racionalizar su uso (1986c, p. 268), acudiendo, por ejemplo, a los expertos relevantes. Si comparamos dos usos del término “artritis” como “se me ha fundido la artritis del salón” y “la artritis que padezco en mis articulaciones desde hace años está afectando a mi muslo” parece que las observaciones de Burge cobran plausibilidad. No reconoceríamos una mínima competencia en el uso del término “artritis” en el primer caso pero sí en el segundo.

⁸⁵ “Tales argumentos sirven para nociones de carácter observacional y teórico, para percepciones y conceptos, para nociones de género natural y de género no natural, para nociones que constituyen un patrimonio especial de los expertos, así como para aquello que en la literatura psicológica se conoce como ‘categorías básicas’. De hecho, creo que, como mínimo, se pueden ofrecer argumentos similares que sirvan para cualquier noción que se aplique a clases públicas de propiedades de los objetos o que se aplique a eventos que se conocen típicamente por medios empíricos” (Burge 1986a, p. 6).

disciplinas, técnicas u oficios en cuyo seno se sitúa el origen del término y/o se determinan las propiedades de los objetos o individuos a los que se debe aplicar la expresión.⁸⁶

Llegados a este punto, podemos recapitular las propuestas de Burge. Para este autor, nuestras prácticas ordinarias de atribución de contenidos mentales dependen de factores de carácter social, externos al hablante, que determinan qué es lo que este hablante cree, desea, teme o espera. En concreto, Burge apela a las intuiciones que respaldan dichas prácticas para mostrar que un sujeto que permaneciera internamente idéntico tendría creencias distintas si el uso socialmente establecido de los términos con los que expresa sus creencias variara y, además, lo hiciera de forma que esta variación del uso le pasara inadvertida al hablante. La identidad de los elementos internos de dos sujetos (aquellos que pertenecen a su experiencia interna y respecto a los cuales concedemos que esos sujetos tienen pleno acceso epistémico) no es condición suficiente para establecer la identidad de sus contenidos psicológicos.⁸⁷ Como vimos en el caso de Bert, dos hablantes pueden asociar (internamente) creencias distintas con un determinado concepto sin que ello suponga que el concepto relevante sea distinto para ambos hablantes. Así, Bert y su doctor diferían internamente en este sentido, al tener el segundo un mayor número de creencias sobre el concepto de artritis y, sin embargo, aceptamos que ese mismo concepto es el que figura en las creencias de ambos. Por otro lado, Bert permanece idéntico en la situación actual y en la situación contrafáctica, lo cual no impide establecer una diferencia en sus contenidos mentales. Mientras que en la situación actual le atribuimos creencias sobre la artritis, en la situación contrafáctica la enfermedad que Bert denomina de ese modo es una enfermedad distinta y, por tanto, la “artritis” de la que hablan los habitantes de la comunidad lingüística contrafáctica, él incluido, no es artritis. El enfoque anti-individualista por el que aboga Burge incide en el hecho de que el uso de las expresiones que tiene un hablante en su vocabulario – un uso determinado por los expertos relevantes de la sociedad a la que pertenece – juega un papel fundamental a la hora de establecer el contenido de los pensamientos de los hablantes. Sólo si tomamos en consideración los usos socialmente asociados a dicha

⁸⁶ Los pastores y los *grafiteros* son tan expertos – en sus respectivos ámbitos – como los médicos, ingenieros o físicos, dado que la aplicación correcta de términos generales como “zalea” o “tageo” viene establecida por personas especialistas en su uso (pastores y *grafiteros*, respectivamente), tan especialistas como lo son los médicos, ingenieros o físicos con respecto a las expresiones “artritis”, “turbina” o “super cuerda”.

⁸⁷ Cf. Loar (1985), p. 181.

expresión, usos que pueden ser parcialmente desconocidos para el sujeto, podremos dar cuenta de manera apropiada de la identidad de sus contenidos psicológicos.

5. Conclusiones

A lo largo de este capítulo hemos tenido ocasión de examinar una serie de propuestas relativas a la semántica de nuestras expresiones lingüísticas cuyas consecuencias afectan a nuestras prácticas de atribución de contenidos mentales. Comprobamos, según se vio, que los defensores de la Teoría de la Referencia Directa (variante de las Nuevas Teorías de la Referencia que se ocupa del valor semántico de nuestros términos) proponen un acercamiento de tipo externista al estudio de la contribución proposicional de ciertos términos singulares y, entre ellos, los nombres propios. Así, las expresiones directamente referenciales no aportan un sentido o una estructura predicativa de carácter general a la proposición que expresan aquellas oraciones en las que figuran, como defendía la semántica Clásica, sino el individuo que tales expresiones designan. Por otro lado, la vertiente Milliana de las Nuevas Teorías de la Referencia pretende hacer explícitos los defectos de la Teoría internista Clásica a la hora de dar cuenta de la referencia de nuestros términos, especialmente los nombres propios y los términos de género natural. A este respecto, Kripke proponía un “cuadro” alternativo para explicar la conexión entre esas expresiones y los aspectos del mundo que designan en función de cadenas causales de comunicación que se sitúan “fuera de las cabezas” de los hablantes. Los experimentos mentales de Putnam y Burge, por su parte, pretenden mostrar de qué manera los factores externos a los hablantes afectan a la semántica de los términos de género natural (en Putnam) y a los usos lingüísticos normativos de los términos generales en su conjunto (en Burge). Ambos concluyen que la identidad de los contenidos mentales que los hablantes expresan haciendo uso de estos términos es dependiente, en parte, de los elementos externos, *de carácter natural o social*. No obstante, es necesario aclarar que, aunque el externismo psicológico focalice su atención en la necesidad de contar con ciertos factores externos a la hora de establecer las condiciones de identidad del contenido psicológico, no por ello las condiciones de identidad son todas ellas de carácter externo.⁸⁸

⁸⁸ Podemos emplear la noción metafísica de “sobreveniencia” que hemos caracterizado en la introducción de nuestro trabajo para aportar claridad en este punto. Los partidarios del internismo mantienen que el contenido mental de un individuo sobreviene a sus propiedades internas. Por tanto, dos individuos que

Por otro lado, es preciso decir que para Burge las relaciones entre nuestros conceptos y los factores externos que participan en su individuación “no tienen que ser tan simples como los ejemplos sugieren [sino que p]ueden ser bastante complejas e indirectas” (Burge 2007, p. 3). Según este autor, por ejemplo, no se requiere la existencia de H₂O en el entorno de un sujeto para que dicho sujeto sea capaz de tener pensamientos acerca del agua. Bastaría con que ese hablante tuviera los conocimientos de química suficientes para formar el concepto de H₂O (Burge 1982, p. 96).

Sea como fuere, hemos visto que dos tesis concernientes a la semántica de nuestros términos lingüísticos presentan claras repercusiones en el ámbito de las atribuciones mentales. El común denominador de estas propuestas y de sus versiones en psicología incide en la necesidad de tomar en consideración la contribución del entorno externo a los hablantes para ofrecer una teoría apropiada de la semántica de sus expresiones y de la identidad de sus contenidos mentales. Además, hemos explicado que la calificación de “interno” o “externo” para una condición de identidad tiene que ver con el hecho de que un hablante pueda tener acceso epistémico inmediato (o no) a dicha condición. En relación con esto, entendimos que en los experimentos mentales de Putnam y Burge se requería la presencia de una premisa fundamental, a saber, que la diferencia entre los entornos permaneciera inadvertida a los sujetos que protagonizaban dichos experimentos. La fuerza de estos experimentos mentales estriba en que aportan plausibilidad a una atribución de contenidos que va más allá de la fenomenología de los hablantes que reciben esas atribuciones, en el sentido en que dos sujetos idénticos en fenomenología pueden merecer atribuciones de distinto contenido. Sin embargo, como veremos a lo largo de los siguientes capítulos, la propuesta externista de contar con factores que pueden superar el alcance cognitivo de los hablantes a la hora de individuar sus contenidos mentales constituye, según muchos autores, el talón de Aquiles de esta teoría.

fueran internamente idénticos en el sentido relevante no podrían poseer contenidos mentales distintos. Para el externista, la base de sobrevenida de los contenidos psicológicos es más amplia y contiene, además de factores internos, ciertas propiedades externas al individuo que posee los contenidos. Por ello, un externista mantiene que es perfectamente posible que dos individuos intrínsecamente idénticos reciban atribuciones con distintos contenidos.

Capítulo 2: El conflicto *prima facie* entre el externismo y el acceso privilegiado y el trilema de Bilgrami.

Un doble objetivo nos guiaba en el capítulo anterior. Primeramente, nos interesamos por los fundamentos semánticos del externismo psicológico, para proponer, en segundo lugar, una caracterización de esta tesis a partir del examen de algunos de los experimentos mentales diseñados para aportar plausibilidad a la individuación externista de los contenidos mentales.

En el presente capítulo pretendemos hacernos cargo de la presunta incompatibilidad entre el externismo psicológico, una tesis metafísica, y una tesis epistémica, relativa al carácter “privilegiado” del conocimiento que un hablante tiene de sus propios contenidos mentales. La orientación metafísica de la discusión emprendida por los internistas y externistas en Filosofía de la Mente es algo que quedó justificado en la introducción. Según vimos, el objeto de esta disputa es la *naturaleza* o *identidad* (individuación) del contenido de nuestros pensamientos. Los internistas, se dijo, reservan una localización exclusivamente interna para aquellos elementos que determinan que nuestros contenidos mentales sean los que son (y no otros distintos) y, por tanto, los factores que debemos tomar en consideración a la hora de establecer la identidad de un pensamiento en tanto que pensamiento de que *P* (y no de que *Q* o *R*) han de ser todos ellos factores internos a nuestras mentes. Los externistas, por su parte, responden que en ocasiones es preciso contar con algunos elementos de carácter externo al hablante para identificar sus contenidos psicológicos, ya sean elementos pertenecientes al entorno natural, como los géneros naturales presentes en dicho entorno, o elementos de carácter social, como los usos lingüísticos vigentes en la comunidad en la que está inserto el hablante.

La descripción de los experimentos mentales que llevamos a cabo en el capítulo anterior nos ha permitido averiguar en qué consiste la internalidad/externalidad de las condiciones de identidad de los contenidos mentales de un sujeto. Si un factor determinante del contenido es externo a un agente esto significa que dicho agente puede no tener conocimiento de dicho factor. Recordemos que Óscar₁ y Óscar₂, los protagonistas del experimento mental de Putnam de (1975c), son dos hablantes enteramente idénticos por lo que respecta a sus *características intrínsecas*. El rasgo del

entorno externo que resulta crucial para establecer una diferencia en sus respectivos estados mentales consiste en la distinta composición química de los líquidos que ambos denominan “agua”. Si, por hipótesis, Óscar₁ y Óscar₂ son ignorantes en cuestiones de química, estos hablantes desconocerán también qué es aquello que determina el distinto carácter de sus creencias. En cuanto al experimento mental de Burge de (1979), comprobamos que su protagonista, Bert, está dispuesto a proferir la oración “creo que sufro de artritis en mi muslo” tanto en la situación actual como en la situación contrafáctica. Ahora bien, en el escenario actual este hablante ignora, en parte, cuál es el uso socialmente establecido para dicha expresión, un uso que restringe la aplicación del término “artritis” a las enfermedades de las articulaciones. Debido a esto, Bert considera (erróneamente) que su aplicación del término “artritis” a una dolencia muscular es correcta. Este hablante, que *permanece invariante en sus aspectos intrínsecos*, también está dispuesto a usar esa expresión de la misma manera en la situación contrafáctica, dado que no hay ninguna diferencia *por lo que a él respecta* entre esas dos situaciones. Sin embargo, en ese contexto negábamos que Bert poseyera el concepto de artritis. La razón es que los usos lingüísticos que la comunidad contrafáctica fija para dicho término son lo suficientemente distintos de los usos actuales como para sostener que la enfermedad llamada “artritis” en esa comunidad no es artritis. El factor que nos compele a decir que Bert posee el concepto de artritis en la situación actual y que nos invita a negar que Bert posea dicho concepto en la situación contrafáctica es un factor que este hablante desconoce, a saber, la diferencia en las prácticas lingüísticas socialmente establecidas para el uso del término “artritis”. Así pues, los experimentos mentales de Putnam y Burge pretenden demostrar que la identidad intrínseca de dos sujetos – identidad que determina que ambos tengan las mismas experiencias subjetivas – no es suficiente para establecer la identidad de sus pensamientos relevantes. La consecuencia de esta afirmación es que un sujeto podría desconocer algunas de las condiciones de identidad de sus contenidos psicológicos (aquellas condiciones que son externas a él) y, en consecuencia, podría tener la creencia de que *P* aunque ignore qué hace que su creencia tenga *P* como contenido, en vez de *Q* o *R*.

Este corolario del externismo psicológico parece entrar en conflicto, como dijimos al comienzo de este capítulo, con una tesis de carácter epistémico, a saber, la tesis de que un hablante posee un *acceso privilegiado* a sus propios contenidos mentales. Recordemos que en la introducción situamos entre las marcas pre-teóricas de los contenidos mentales la presuposición de que, por lo general, conocemos nuestros

propios pensamientos y que, además, los conocemos con mas seguridad y fiabilidad que un tercero que trate de establecer informes sobre nuestras creencias, deseos, temores, etc. Ese privilegio epistémico podría verse en entredicho si, atendiendo al externismo, la identidad de nuestros contenidos depende de factores que desconocemos. Dicho de otra manera, si puedo ignorar qué es lo que determina que mis contenidos mentales sean los que son, ¿podría seguir sabiendo que *esos* (y no otros) son mis pensamientos? Esa es la pregunta que nos guía en este trabajo y todos nuestros esfuerzos irán dirigidos a evaluar los distintos planteamientos que intentan responder a esta cuestión, ya sea de manera afirmativa o negativa. Por el momento, dedicaremos este capítulo y el siguiente al *problema del logro* (“the achievement problem”) del conocimiento privilegiado de nuestros contenidos mentales desde un marco externista acerca de lo mental.⁸⁹

En las próximas secciones, profundizaremos en la comprensión de la accesibilidad privilegiada a nuestros propios pensamientos para presentar, a continuación, algunas propuestas destinadas a demostrar que la individuación externista no pone en peligro el carácter especial del conocimiento de nuestros contenidos mentales. Asimismo, examinaremos el argumento desarrollado por A. Bilgrami en forma de trilema para poner en evidencia la supuesta incompatibilidad entre, por un lado, el externismo psicológico originado en las ideas de Putnam y de Burge y, por otro lado, el conocimiento privilegiado de nuestros propios pensamientos.

1. *Las líneas fundamentales del conflicto prima facie*

En esta primera sección queremos explicar de la manera más precisa posible en qué sentido el externismo psicológico y el acceso privilegiado parecen ser dos tesis difícilmente conjugables. Para ello conviene aclarar, primero, en qué consiste el acceso privilegiado a los propios pensamientos, qué rasgos muestra dicho acceso y cuál es su alcance. A continuación, veremos de qué manera la filosofía ha dado cuenta tradicionalmente del conocimiento introspectivo y, para finalizar, expondremos los motivos por los que la individuación externista de los contenidos mentales parece obstaculizar dicho conocimiento.

Tal y como hemos señalado, nuestras intuiciones sobre los contenidos mentales nos invitan a considerar que éstos son objeto de un conocimiento “especial” o “privilegiado” por parte del hablante que las posee. Si Juan afirma a menudo que los

⁸⁹ Cf. Brown (2009), p. 767 y Davies (2000), p. 391.

franceses no son dignos de confianza es natural suponer que este hablante tiene una creencia al respecto y que, además, sabe que tiene dicha creencia. Desde luego, nosotros también podemos saberlo a partir de sus declaraciones y, en esta medida, seremos capaces de elaborar informes como:

(1) Juan cree que los franceses no son dignos de confianza.

No obstante, presumimos que existe una intuitiva diferencia entre *nuestro* conocimiento de que Juan posee una creencia de esa índole y el conocimiento que *Juan mismo* tiene de su propia creencia. Para empezar, puede que este sujeto esté intentando engañarnos, o quizá se trata de alguien dotado de un peculiar sentido del humor que suele decir, burlonamente, que los franceses no son gente de fiar cuando está en presencia de su mujer, natural de Francia. Asimismo, somos capaces de imaginar una situación en la que Juan habría tenido esta creencia sin que nunca hubiera llegado a pronunciarse al respecto, de manera que nadie habría podido afirmar (1). Sin embargo, suponemos que en ese escenario hipotético el propio Juan sí habría sabido que cree que los franceses no son dignos de confianza. Podemos apreciar, con ello, que nosotros necesitamos observar la conducta de Juan y entender sus declaraciones para inferir (de modo perfectamente falible) qué es lo que cree, mientras que él no requiere esta clase de indicios empíricos para conocer sus propias creencias. Suponemos que Juan posee un cierto “privilegio” en su *conocimiento de primera persona*, esto es, que este hablante es capaz de conocer de una manera especialmente fácil y fiable qué es lo que piensa, mientras que nosotros carecemos de esas facilidades para elaborar informes acerca de sus contenidos mentales.

Contemplado desde un punto de vista subjetivo, parece plausible considerar que *el auto-conocimiento*, esto es, el conocimiento que tenemos de nuestros propios pensamientos, es un conocimiento dotado de *autoridad*, *inmediatez* y *notoriedad* (Wright/Smith/MacDonald 2000, p. 2). Por lo que respecta a la *autoridad*, creemos que, en general, nadie sabe mejor que nosotros mismos qué es lo que pensamos. Desde luego, podríamos cometer un *lapsus* al expresarnos y decir, por ejemplo, “creo que este bolígrafo se ha quedado sin punta” cuando en realidad queríamos referirnos a un lápiz, pero parece que siempre cabría alegar que sabíamos lo que pensábamos aunque nos equivocamos al comunicarlo a los demás. Para comprender la naturaleza de la *inmediatez* del auto-conocimiento debemos comparar el proceso que seguimos para atribuir un pensamiento a un tercero y el que emprendemos para decir qué es lo que pensamos nosotros mismos o, como diremos a partir de ahora, para realizar *una auto-*

atribución. Por lo general, afirmar que un sujeto mantiene un pensamiento requiere llevar a cabo inferencias a partir de su conducta verbal y no verbal: sus declaraciones, sus acciones y reacciones frente a las declaraciones y la conducta de otros hablantes nos proporcionan las evidencias que justificarán nuestros informes. Sin embargo, habitualmente no necesito escuchar mis palabras ni observar mi comportamiento para saber lo que yo misma pienso.⁹⁰ En la mayoría de los casos las auto-atribuciones pueden establecerse sin necesidad de colegirlas a partir de observaciones empíricas. Por último, en lo que se refiere a la *notoriedad*, cabe pensar que nuestras creencias, deseos, esperanzas y temores son estados mentales que se nos manifiestan. A pesar de que no necesito tener consciencia en el momento presente de todos y cada uno de mis pensamientos (Searle 1983, p. 2), usualmente se presupone que un sujeto que declara temer que *P* o desear que *Q* tiene consciencia de ello en el momento de realizar la declaración.

La idea de que el conocimiento que tenemos de nuestros propios estados mentales de carácter proposicional es diferente del conocimiento que poseemos de los pensamientos ajenos ha recibido el nombre de *asimetría entre la primera y la tercera persona*. Como vemos, dicha asimetría consiste en que el conocimiento de nuestros propios pensamientos goza de una especial autoridad, fundamentada en la inmediatez del acceso introspectivo y la notoriedad de los pensamientos así captados.

Por otro lado, la asimetría entre la primera y la tercera persona se revela como un *factum*, como un dato bien confirmado por nuestra concepción pre-teórica en lo que respecta no sólo a los estados mentales de carácter proposicional, sino también a los demás ámbitos de nuestra vida psicológica. En este sentido, suele señalarse que la actividad mental consciente comprende tanto estados psicológicos dotados de contenido como *aspectos fenoménicos o cualitativos*.⁹¹ Aunque los primeros son el objeto principal de nuestro estudio – en la medida en que las tesis externistas se aplican a la individuación de esa clase de estados psicológicos –, a continuación caracterizaremos brevemente los segundos, las experiencias sensoriales, que constituyen la faceta fenomenológica o cualitativa de nuestra vida mental. Así, en (1974) Nagel acuñó una denominación muy extendida para el campo de estados fenoménicos, a saber, “*lo que se*

⁹⁰ “Infiero de la observación de su conducta que debe ir al médico; pero esta conjetura no la hago, para mí mismo, sobre la base de la observación de mi conducta. O, mejor: también hago esto algunas veces, pero no en casos análogos” (Wittgenstein 1967, p. 94).

⁹¹ Cf. Moya (2004), p. 18 y Chalmers (1996), pp. 11 y ss.

siente al ser (como) un ser consciente".⁹² Según este autor, un organismo posee una vida mental consciente sólo si existe algo que es "el sentirse como" ese organismo. Como es natural, el tipo de órganos sensoriales que posee un ser vivo determina el carácter de sus experiencias y, en consecuencia, "sentirse como" un ser humano sin discapacidades sensoriales es distinto de "sentirse como" un ser humano ciego y, desde luego, es radicalmente diferente de "sentirse como" un murciélago. A partir de esto, podemos decir que la asimetría mencionada se presenta en el conocimiento de ambos aspectos de nuestra vida mental, tanto de los estados mentales con contenido como de las experiencias subjetivas conscientes. Sabemos más sobre nuestra vida psicológica, considerada en un sentido general, que sobre la vida psicológica de los demás y lo que sabemos sobre la nuestra lo podemos aseverar con mayor seguridad que nadie. Por esta razón, en muchos textos de introducción a la Filosofía de la Mente suele mencionarse la accesibilidad privilegiada como una de las "marcas" que caracterizan el objeto de esta disciplina, ante las evidentes dificultades para elaborar una caracterización de la mente no comprometida con alguna concepción particular de la misma.⁹³ La importancia del acceso privilegiado a los propios contenidos y experiencias a la hora de delimitar el ámbito y características de lo mental es tan destacada que algunos autores han llegado a sostener que el privilegio epistémico de la introspección es la única característica genuina del ámbito psíquico:

"Mi propuesta es que el reino de lo mental no es nada más que el objeto de la capacidad cognitiva que me dota de un acceso especial: esto es, el área de lo que es conocido por mí de una manera tal que no es conocido del mismo modo por nadie más" (Farkas 2008, p. 22).

No queremos discutir aquí si la accesibilidad privilegiada constituye el único o mejor modo de demarcar el terreno de lo mental. Nos limitaremos a asumir que el conocimiento privilegiado es una característica importante de nuestra vida psicológica que se extiende a todos sus ámbitos.

En lo que sigue, sin embargo, querríamos señalar algunas diferencias entre el acceso introspectivo a nuestros propios contenidos mentales y el acceso que se da con respecto a nuestras experiencias de tipo cualitativo. En este sentido, aunque anteriormente se indicaron cuáles son los rasgos del privilegio que existe en el

⁹² Aunque la traducción literal de la expresión usada por Nagel, "what is like to be", es "lo que es ser como", proponemos al respecto "lo que se siente al ser como" por ser más natural en castellano.

⁹³ Cf. Kim (1998), pp. 16 y ss.

conocimiento de nuestros propios contenidos (autoridad, inmediatez, notoriedad), ahora queremos precisar aún más el carácter de dicho privilegio, comparándolo con aquél que se presenta en nuestro acceso al aspecto fenomenológico de nuestra vida consciente. Atendiendo a las observaciones de Wright (1998 pp. 14 y ss.), es conveniente distinguir las características y condiciones de las *aserciones fenoménicas*, relativas a nuestras sensaciones y experiencias cualitativas, de las características y condiciones que corresponden a las *aserciones actitudinales*, consideradas en tanto que informes sobre nuestros estados mentales dotados de contenido. Para empezar, la autoridad de la que gozamos al realizar aserciones fenoménicas como “me duele una muela” o “me pica” es una *autoridad fuerte*. Es difícil no dar crédito a nuestro interlocutor cuando profiere este tipo de aserciones sobre sí mismo y, recíprocamente, esperamos que a los demás les baste con nuestra afirmación de que tenemos un dolor para que crean en la verdad de nuestro informe. Además, consideramos que las aserciones fenoménicas son *transparentes* al sujeto y, por ello, no resulta demasiado sensato proferir con seriedad una aserción como “me pica una mano pero no estoy seguro”. Por último, las aserciones fenoménicas son afirmaciones *carentes de fundamento*, en el sentido de que el testimonio de un hablante que realiza aserciones de esta clase suele ser aceptado sin necesidad de justificaciones adicionales. Por ejemplo, cuando alguien asegura que se le han dormido las piernas, ningún hablante que entienda sus palabras le haría preguntas como “¿qué razón tienes para saber tal cosa?” o “¿en qué te basas para asegurarlo?”.

No obstante, Wright indica que algunas de las aserciones que toman por objeto estados mentales portadores de contenido carecen de los rasgos que reconocemos en las aserciones fenoménicas. En concreto, este autor se refiere a las aserciones actitudinales que son resultado de *un proceso de auto-interpretación del sujeto*. Por ejemplo, Juana puede haber necesitado cierta reflexión para ser capaz de confesar “deseo que él me quiera” refiriéndose al actual marido de su hermana, lo cual pone de manifiesto que esa actitud psicológica no le resultaba completamente transparente en un determinado momento. Además, si fuera interpelada por su interlocutor mediante una pregunta como “¿realmente piensas que ese es tu deseo?” no habría nada de inusual en ello.⁹⁴ Parece que, en estas ocasiones, cabe exigir a los hablantes una justificación sobre sus auto-atruciones, lo que nos lleva a considerar que en los casos de auto-interpretación la autoridad de los hablantes para informar de sus creencias y deseos no es tan fuerte como

⁹⁴ Compárese esa pregunta con otra acerca de una aserción fenoménica como “¿realmente piensas que te duele?”.

en el ámbito de las aserciones fenoménicas. Sin embargo, la existencia de auto-atribuciones que requieren un proceso de auto-interpretación no impide que en los casos más básicos un sujeto posea autoridad para pronunciarse sobre sus propios pensamientos, aunque se trate de una autoridad más débil que la existente en las aserciones fenoménicas (Wright 1998, p. 17). En particular, este autor afirma que la autoridad que se presupone en el conocimiento que un agente tiene de sus propios pensamientos no es más que la presunta aceptabilidad de su testimonio al respecto. Esto supone que no podemos despojar a un sujeto en todos los casos de su papel de informador especialmente fiable acerca de sus contenidos mentales.

La consideración de la asimetría por parte de algunos autores de la Filosofía Moderna, orientada a explicar la autoridad, inmediatez y notoriedad que caracterizan el auto-conocimiento, se ha basado en una concepción de éste que sigue muy de cerca la descripción del conocimiento que obtenemos de los objetos externos por vía perceptiva. Shoemaker, en un magnífico estudio sobre el *modelo perceptivo del auto-conocimiento* (1994) afirma que la filosofía tradicional (especialmente, Descartes y los empiristas ingleses) ha hecho uso de un modelo teórico que explica la introspección en tanto que ejercicio de un cierto “sentido interno” capaz de percibir “objetos” mentales, comparable al ejercicio de nuestro sentido de la vista en la percepción de objetos externos. Sin embargo, Shoemaker dedica el grueso de su trabajo a señalar algunas de las dificultades que presenta el empleo de esta metáfora, cuestionando que los elementos que están presentes en la percepción visual también lo estén en la “percepción” introspectiva. La percepción visual parte de la existencia de un sentido, el sentido de la vista, cuyo ejercicio nos permite obtener creencias acerca de hechos y objetos externos. Esto implica hacernos conscientes de su existencia y propiedades a partir de aquellas impresiones y apariencias, distintas de los hechos y objetos pero causadas por éstos, hacia los que se dirige nuestra atención. Este modelo no puede ser trasladado sin pérdida a la “percepción” introspectiva. Por ejemplo, la existencia y naturaleza de ese presunto sentido interno encargado de obtener conocimiento introspectivo queda en la más absoluta oscuridad. Tampoco resulta plausible suponer que los pensamientos “causen” nuestra propia conciencia de los mismos. Además, no encontramos “apariencias” de pensamientos y experiencias a partir de las cuales accedamos mediatamente a los pensamientos y sensaciones y, desde luego, la ausencia

de apariencias en la introspección impide que en este ámbito del conocimiento se den aquellos errores propios de la percepción visual.⁹⁵

De lo dicho hasta aquí podemos extraer la siguiente conclusión. A pesar de que el análisis moderno del *factum* de la asimetría es, como vemos, poco satisfactorio, la idea de que poseemos un acceso privilegiado en mayor o menor medida a los elementos que constituyen nuestra vida mental, incluyendo los estados psicológicos de carácter proposicional, encuentra un firme respaldo en nuestras intuiciones. De esta manera, parece que si una teoría filosófica entrara en conflicto con el acceso privilegiado, dicho conflicto redundaría en perjuicio de la teoría en cuestión, puesto que, a primera vista, estamos poco dispuestos a renunciar al privilegio epistémico de primera persona.

Este es el momento de acercarnos con más atención a las razones del presunto conflicto entre el externismo y el acceso privilegiado. La individuación externista de los contenidos mentales supone que la identidad de las creencias y deseos de un hablante depende esencialmente de rasgos naturales o sociales de su entorno. Dichos rasgos se encuentran fuera de los límites de la mente de su mente y, por tanto, fuera de los límites del acceso privilegiado que presuponemos en este agente. Aquello de lo que depende que las creencias de Óscar₁ tengan el carácter que tienen, es decir, que sean creencias sobre el agua y no sobre el “agua gemela” o cualquier otro líquido, es algo que se encuentra “fuera de su cabeza”. La situación externa de algunas de las condiciones de identidad del contenido un pensamiento implica que, como en el caso del propio Óscar₁, un sujeto puede ignorar o desconocer en parte qué es aquello que determina que su pensamiento sea el que es.⁹⁶ La consecuencia que obtenemos es que algunos de los

⁹⁵ Podemos encontrar unas observaciones muy similares a las de Shoemaker en el texto clásico de G. Ryle de (1949). Según este autor “la doctrina oficial” acerca de la naturaleza de la mente describe el acceso introspectivo como una “mirada” a nuestro interior. El carácter especial de esta mirada permite que un individuo siempre sea capaz de saber de manera infalible qué siente o piensa y, además, que pueda saberlo de una forma tal que ninguna otra persona sería capaz de conocerlo del mismo modo. Los pensamientos que otros me atribuyen son resultado de “inferencias problemáticas” (Ryle 1949, p. 18), puesto que mis operaciones mentales, mis experiencias y pensamientos permanecen por principio ocultos a los demás. Sin embargo, aunque esta doctrina explica (aun de manera metafórica) por qué disfrutamos de un acceso privilegiado a los eventos que suceden en nuestra mente, lo hace al precio de impedir que los demás puedan acceder a los mismos, dado que la “doctrina oficial” otorga a tales eventos el estatus de objetos privados. De esto se desprende que un partidario de la doctrina en cuestión debe hacer frente a una consecuencia muy problemática de sus tesis, a saber, “que no tiene buenas razones para pensar que existen otras mentes además de la suya” (Ryle 1949, *ibídem*).

⁹⁶ No está de más señalar que las tesis externistas no implican que un individuo siempre desconoce todos aquellos factores externos que condicionan la identidad de los contenidos de sus pensamientos. Que un determinante del contenido sea externo implica que *puede* ser desconocido al sujeto, no que *necesariamente* lo sea. Bastaría con que Óscar₁ y Óscar₂ tuvieran conocimiento de la diferencia existente entre los géneros naturales que uno y otro denominan “agua” para mostrar que el carácter externo de un factor es compatible con el hecho de que un sujeto lo conozca (empíricamente).

factores que determinan parcialmente algunos de nuestros pensamientos pueden ser factores que se sitúen fuera del alcance de nuestro acceso privilegiado.

¿En qué sentido podría existir un conflicto entre el externismo psicológico y los tres rasgos (inmediatez, notoriedad, autoridad) que caracterizan el auto-conocimiento de nuestros pensamientos? De acuerdo con el externismo, que una creencia sea la creencia de que *P* (y no de que *Q* o *R*) puede depender, en parte, de qué géneros naturales se encuentran presentes en el entorno del hablante que posee dicha creencia, así como de los usos que otros miembros de su comunidad lingüística determinan para las expresiones que él utiliza. Estos factores externos, cuyo conocimiento estaría sujeto a indagaciones empíricas, son tales que el hablante en cuestión podría ignorarlos total o parcialmente. Sin embargo, esto supone que un hablante debería emprender investigaciones empíricas sobre los géneros naturales o los usos lingüísticos de otros hablantes para poder averiguar que su creencia es la creencia de que *P* (y no de que *Q* o *R*). Si del externismo se desprende que conocer el contenido de nuestros pensamientos requiere, en primer lugar, conocer hechos empíricos, entonces el externismo parece atentar contra la *inmediatez* del auto-conocimiento. Por otra parte, la *notoriedad* del conocimiento en primera persona también parece resultar dañada por la individuación externista de los pensamientos. Por hipótesis, dos hablantes podrían tener creencias distintas (como Óscar₁ y Óscar₂) sin advertir la diferencia en sus estados mentales, al situarse el fundamento de la misma fuera del ámbito de lo que les resulta manifiesto. Además, la *autoridad* que se presume en nosotros a la hora de conocer mejor que ninguna otra persona cuáles son nuestros pensamientos también se vería en peligro, puesto que un hablante que conociera mejor que yo qué factores determinan el contenido de mis creencias podría conocer mejor que yo qué contenido tienen mis creencias.⁹⁷ Consiguientemente, el externismo parece atentar contra los tres rasgos de la accesibilidad privilegiada.

Como veremos a lo largo de este trabajo, muchos autores han presentado una serie de planteamientos que, con distinto grado de sofisticación, pretenden poner de manifiesto las fricciones existentes entre la individuación externista de los contenidos y el acceso privilegiado a los mismos. En lo que sigue, nos referiremos a la enunciación del problema que acabamos de realizar como *planteamiento simple* o *enunciación del*

⁹⁷ “Dado que la relación externa no se determina subjetivamente, el sujeto no tiene autoridad sobre ello. Una tercera persona podría muy bien estar en mejor posición que el sujeto para saber sobre qué objeto está pensando el sujeto y por ello, estar en mejor posición de saber cuál era el pensamiento” (Woodfield 1984, p. viii)

conflicto prima facie entre el externismo y el acceso privilegiado. Conviene hacer una segunda aclaración terminológica. Hablaremos de autores *incompatibilistas* para distinguir a aquellos que mantienen la existencia de un conflicto, del tipo que fuere, entre el externismo y el auto-conocimiento, mientras que calificaremos de *compatibilistas* a aquellos autores que rechazan la existencia de un problema de esa índole. En lo que sigue, presentaremos las propuestas compatibilistas de Burge y Davidson frente al planteamiento simple, así como la original reelaboración del mismo en forma de trilema por parte de Bilgrami.

2. *La propuesta compatibilista de Burge: el modelo basado en la inclusión*

Dedicaremos esta sección a examinar la propuesta de Burge para mantener, simultáneamente, que los contenidos de nuestros pensamientos reciben una individuación externista y que el conocimiento que tenemos de los mismos es tan directo, fiable e inmediato como nos dicen nuestras intuiciones al respecto. El programa compatibilista de este autor, cuyo famoso argumento a favor del anti-individualismo psicológico expusimos en el capítulo anterior, ha recibido muchos apoyos de la comunidad externista. Además, como veremos, algunas observaciones que figuran en los trabajos que Burge presentó en la década de los 80 del pasado siglo son tan sugerentes que sentaron las bases de polémicas que se extendieron a lo largo de la década siguiente.

Comencemos señalando que la estrategia compatibilista que aporta Burge se despliega en dos frentes. En primer lugar, pretende impedir que los partidarios del individualismo psicológico tomen el acceso privilegiado a los propios pensamientos como un dato favorable a sus propias tesis. En segundo lugar, Burge elabora un modelo teórico para explicar la especial naturaleza de las relaciones epistémicas entre un sujeto y sus propios pensamientos. Esta doble tarea tiene como fin desligar dos cuestiones: por una parte, la especial autoridad del conocimiento de los propios contenidos mentales y, por otra, las tesis (externistas o no) sobre la naturaleza de dichos contenidos. De este modo, Burge trata de demostrar que un defensor del externismo no se encuentra en desventaja frente al partidario del internismo a la hora de respetar el *factum* del acceso privilegiado.

Examinemos, pues, el argumento propuesto por Burge para contrarrestar el intento de emplear el privilegio epistémico de primera persona en provecho del

internismo. Este autor ha señalado en varias ocasiones que dicho intento tiene que ver con una interpretación ilegítima de la brecha o desfase (“gap”) causal existente entre el mundo externo y nuestros pensamientos acerca de él (Burge 1988a, p. 650 y 1986b, p. 195). Además, en sus textos suele destacar la maestría de Descartes para poner de manifiesto la existencia de esa brecha en la primera de sus *Meditaciones Metafísicas* (1641, pp. 93-97) y, muy especialmente, en la famosa “hipótesis del Genio Maligno”. Con el objetivo de entender en qué consiste este desfase causal, y qué tiene que ver con el asunto que nos interesa, parece conveniente acercarnos al texto cartesiano de (1641) y al lugar que en él ocupa la hipótesis escéptica aludida. El objetivo de esta obra es desarrollar un proyecto de fundamentación del conocimiento y el modo en que Descartes intenta abordar una investigación de este calado consiste en enfrentarse a una serie de hipótesis que, de darse, pondrían en cuestión las bases más elementales de todo nuestro saber. Para el autor responder a la pregunta “¿qué nos es dado conocer con absoluta certeza y de qué manera?” requiere descartar previamente una respuesta afirmativa a la cuestión “¿es posible que existan motivos para dudar incluso de lo que nos parece completamente cierto?”. Una teoría del conocimiento como la cartesiana, que tome a éste como conocimiento de certezas absolutas (Rábade 1971, p. 25), sólo puede elaborarse tras haber demostrado previamente que no hay lugar para una duda radical que lo cuestione todo, incluso lo que nos parece absolutamente cierto. En lo que sigue, nos centraremos en la hipótesis del Genio Maligno, por ser ésta la formulación cartesiana de la duda radical que Burge retoma en su argumento.

Descartes diseña una situación en la que un agente externo, un poderoso Genio Maligno, es el encargado de inducirle todas sus experiencias y pensamientos. Este Ser provoca en él determinadas sensaciones visuales y táctiles que le invitan a creer que está sentado, en bata, delante del fuego, meditando sobre el origen de nuestro conocimiento. Es decir, en esa situación de engaño las experiencias sensoriales de Descartes serían cualitativamente indistinguibles de aquellas experiencias que tendría él mismo si, en efecto, estuviera sentado, vestido con bata, delante del fuego. Un escenario semejante sirve para plantear provisionalmente la posibilidad de que Descartes esté siendo completamente engañado por lo que se refiere al origen de su vida psíquica. Lo interesante es que Descartes no tendría noticia de las malas artes del Genio Maligno, puesto que, *por lo que a él respecta*, todo sería exactamente igual, cualquiera que fuera el origen causal de sus pensamientos y sensaciones (bien el Genio Maligno, bien los objetos externos como el fuego, la silla y la habitación). La potencia del escenario

cartesiano radica en que, por lo que parece, cualquiera de nosotros podríamos estar ocupando el lugar de su protagonista y, si así fuera, también seríamos incapaces de averiguar desde nuestro punto de vista subjetivo cuál es el origen causal efectivo de nuestra vida interior (Burge 1986b, p. 195). Lo que pretendemos extraer de esta conclusión escéptica es que existe una vinculación entre la brecha causal y el conocimiento privilegiado de lo mental, vinculación que consiste en que conocemos de manera privilegiada nuestras experiencias y cogniciones, mientras que este privilegio no se extiende al conocimiento de las causas (el mundo o el Genio) de tales efectos fenoménicos y cognitivos.

No vamos a detenernos en los recursos que propone Descartes para hacer frente a esta hipótesis y restaurar, con ello, la confianza en el origen objetivo de nuestras experiencias sensoriales. Sólo hemos planteado la hipótesis del Genio Maligno para ver de qué manera un partidario del individualismo psicológico podría sacar provecho de tal escenario. Según esta tesis, recordemos, qué contenido tienen nuestros pensamientos es algo que depende por entero de los elementos internos a la mente de los hablantes, esto es, de elementos que los hablantes son capaces de conocer de manera privilegiada mediante acceso introspectivo. ¿De qué forma la intuición que pone en juego el escenario del Genio Maligno parece constituir un apoyo para las tesis internistas de la individuación de contenidos? Para verlo con más claridad, separemos ambas tesis:

- (A) Vinculación entre el acceso privilegiado y la brecha causal: es concebible que estemos radicalmente equivocados sobre el origen causal de nuestros pensamientos (por ejemplo, podríamos estar equivocados al pensar que dicho origen es el mundo externo más bien que un Genio capaz de provocar en nosotros esos efectos psicológicos).
- (B) Individualismo psicológico: nuestros pensamientos quedan completamente determinados por factores internos a nosotros mismos, es decir, por elementos que pertenecen a nuestra vida psíquica. Esto supone que ninguna variación del entorno externo, por muy radical que fuera, conllevaría un cambio en la identidad de los contenidos de nuestros pensamientos.

El razonamiento individualista a partir de (A) discurre de la siguiente manera. Por (A), conocemos nuestros pensamientos de manera privilegiada, pero ese privilegio no se extiende a nuestro conocimiento de su procedencia causal. Al no conocer de manera privilegiada cuál es el origen efectivo de mis pensamientos actuales, cabría pensar en una situación como la que describe la hipótesis del Genio Maligno, en la que nuestra

vida mental (la misma que tenemos ahora) habría tenido una procedencia causal muy distinta. Si ese hubiera sido el caso, la identidad de mis pensamientos se habría conservado, a pesar de la diferencia en sus causas. Esta clase de escenarios son empleados para aportar plausibilidad a (B), el individualismo sobre lo mental, en cuanto que pretenden reforzar la creencia de que la identidad de nuestros pensamientos es inmune a variaciones extremas en su procedencia externa y, por tanto, dicha identidad depende únicamente de factores internos a los sujetos.

Una vez planteado el argumento individualista a partir del acceso privilegiado, examinaremos la táctica que Burge plantea para defender que el diferente carácter del acceso epistémico a los elementos separados por la brecha no constituye un argumento favorable al individualismo. Para empezar, conviene aclarar qué alcance tiene para este autor el privilegio epistémico de primera persona, si bien no describiremos por el momento los rasgos de dicho privilegio. Según Burge, la tesis (A), que constituye la intuición de la que hace uso Descartes al diseñar el escenario del Genio Maligno, es una tesis esencialmente correcta. Esta intuición tiene su origen, como vimos, en el hecho de que el privilegio epistémico está restringido al conocimiento de los elementos de nuestra vida mental. De esta forma, aun sabiendo bien qué es lo que pensamos *actualmente*, podríamos estar radicalmente equivocados en cuanto a la naturaleza de lo que causa nuestros pensamientos. Ahora bien, la existencia de este desfase causal y de la asimetría epistémica sustentada en él no conlleva *per se* un acceso epistémico privilegiado a los pensamientos que *tendríamos* en una situación contrafáctica como el escenario del Genio Maligno. Dicho de otra manera, Burge no está interesado en plantear una objeción a la tesis (A), restringida a nuestros pensamientos actuales, sino a la interpretación de (A) que pone en juego el partidario del individualismo psicológico (Burge 1986b, p. 196 y 1988b, p. 211). Para este autor, en consecuencia, no es legítimo el paso que individualista da en su argumentación desde (i) sé lo que pienso de manera especial pero lo que pienso podría ser erróneo porque no conozco de manera privilegiada qué es lo que lo causa, hasta (ii) sé que pensaría lo mismo incluso si estuviera equivocado sobre las causas de mi pensamiento. La premisa (i) se apoya en el conocimiento privilegiado de *lo que pienso actualmente*, mientras que la premisa (ii) se sustenta en el conocimiento privilegiado de *lo que pensaría* en caso de que los agentes externos causantes de mis pensamientos variaran. Por tanto, para establecer (ii) deberíamos tener tanto la capacidad de saber privilegiadamente qué pienso en la situación actual, como la capacidad de saber de la misma manera qué pensaría en la

situación contrafáctica y, por descontado, también la capacidad de comparar lo pensado en una situación y en la otra. Sin embargo, para Burge esto supondría extender de manera fraudulenta el ámbito del acceso privilegiado, que se restringe por derecho propio a (i).

Examinado el modo en que Burge desactiva el intento de reforzar el individualismo psicológico haciendo uso del acceso privilegiado y del desfase causal que lo fundamenta, emprenderemos el análisis del segundo frente abierto por este autor para completar su proyecto compatibilista. Como dijimos, la segunda vertiente de su argumentación tiene que ver con la elaboración de un modelo que explique las condiciones de posibilidad del privilegio del conocimiento en primera persona. La tesis fundamental de Burge es que los elementos desplegados en dicho modelo son completamente independientes de la concepción que uno mantenga sobre la individuación del contenido mental.

A este respecto conviene comenzar aclarando qué rasgos tiene el conocimiento de primera persona y, muy especialmente, a qué clase de pensamientos se aplica. Para este autor, la accesibilidad privilegiada es una marca del *auto-conocimiento básico*, esto es, del conocimiento que tenemos de nuestras actividades mentales (“pienso”), de nuestros contenidos actitudinales (“creo *que la silla está rota*”) y de nuestros pensamientos tomados como contenido de otros pensamientos (“creo *que temo una derrota del Rayo Vallecano*”). Burge denomina “juicios de tipo *Cogito*” a los juicios o pensamientos resultantes del proceso de auto-conocimiento básico, es decir, juicios cuyo contenidos son idénticos a nuestros propios contenidos psicológicos. El auto-conocimiento básico se caracteriza por ser *auto-verificativo*, *directo*, y *no requerir evidencias o comparaciones empíricas*.⁹⁸ Con el fin de justificar la presencia de estas tres marcas en el conocimiento de primera persona en los casos más básicos, Burge elabora una descripción del proceso de formación de los juicios de tipo *Cogito* conocido como

⁹⁸ Cabe preguntarse qué relación tienen estos tres rasgos que Burge atribuye al auto-conocimiento básico (ser auto-verificativo, directo y no requerir evidencias empíricas) y los tres rasgos que hemos atribuido al auto-conocimiento en general (inmediatez, notoriedad y autoridad). En este sentido, la notoriedad e inmediatez del auto-conocimiento en general podrían asimilarse, respectivamente, al carácter directo y no-empírico del auto-conocimiento básico. De este modo, si aquello que estamos pensando nos resulta notorio, sabemos directamente qué es lo que pensamos y, de no requerir evidencias empíricas en este sentido, podremos obtener ese (auto-)conocimiento de manera inmediata. Sin embargo, el carácter auto-verificativo del auto-conocimiento básico y la autoridad que poseemos para conocer nuestros pensamientos en general no parecen ser idénticos. En este sentido, mientras que Burge afirma que en los juicios de tipo *Cogito* basta con pensar reflexivamente para tener un pensamiento reflexivo verdadero, en el caso del auto-conocimiento en general pueden darse situaciones puntuales de auto-engañó, aunque se presuponga globalmente una especial autoridad en el sujeto que realiza auto-atruciones de esta clase de pensamientos.

modelo basado en la inclusión (Bernecker 1996) o *modelo de la reutilización* (Peacocke 1996). Según dicho modelo, podemos distinguir en el conocimiento de nuestros pensamientos dos tipos de juicios, a saber, los *juicios de primer orden*, es decir, los pensamientos que conocemos, y los *juicios de segundo orden*, que tienen como contenido los juicios de primer orden. Pongamos como ejemplo el siguiente informe de auto-conocimiento:

(2) Sé que creo que Esther no aprecia la puntualidad.

En (2) podemos distinguir un juicio de primer orden, “creo que Esther no aprecia la puntualidad”, y un juicio de segundo orden que versa sobre el primero: “*Sé que* creo que Esther no aprecia la puntualidad”.⁹⁹ El juicio de primer orden, como vemos, queda incluido en el juicio de segundo orden como su contenido. Por consiguiente, los conceptos que figuran en el juicio de primer orden son, también, los conceptos que figuran en el juicio de segundo orden. La inclusión conceptual permite justificar el carácter *auto-verificativo* de los juicios de tipo *Cogito*, esto es, el hecho de que en los casos más básicos juzgar acerca de lo que pensamos es juzgar siempre algo verdadero. Así, no existe una distancia epistémica entre nuestra creencia y el pensamiento que versa sobre ella, puesto que la creencia juzgada está contenida en el propio pensamiento. La garantía de que una auto-adscripción como (2) tiene un carácter *directo* reside en que su verdad se deriva del mero hecho de ser realizada, *sin que se requiera hacer uso de justificaciones sobre cuestiones empíricas* o sobre supuestos “objetos” introspectivamente accesibles. Esta última observación evidencia que Burge está interesado en marcar una distancia con respecto a la visión tradicional del acceso privilegiado, basada en la percepción de objetos mentales tal y como expusimos en la sección anterior.¹⁰⁰ De hecho, este autor afirma explícitamente que el acceso epistémico al contenido de nuestros propios pensamientos carece del carácter objetivo que tiene nuestro acceso perceptivo a los objetos externos (Burge 1988a, pp. 657 y ss.). Esta *objetividad* se manifiesta en un doble sentido. Por un lado, se entiende que el acceso perceptivo es objetivo en cuanto se dirige hacia *objetos y estados de cosas independientes del conocimiento* mismo. Ver una pared blanca delante de mí y,

⁹⁹ Según las pautas que proporciona Burge no es necesario que el juicio de primer orden sea una actitud proposicional completa como en “sé que *creo que Esther no aprecia la puntualidad*”. Podríamos haber puesto como ejemplo “Sé que pienso”, donde “pienso” sería un juicio de primer orden.

¹⁰⁰ Como tendremos ocasión de ver a lo largo de nuestro trabajo, la invitación a dejar de lado esa visión es una constante en los textos de los externistas. Así, en Burge (1988a), p. 660; Heil (1988), pp. 248 y ss.; Davidson (1987), pp. 67 y ss.; Falvey y Owens (1994), p. 117 y Putnam en Pessin/Goldberg (1996), p. xviii.

consecuentemente, saber que hay una pared blanca delante de mí es un conocimiento sobre la pared misma, cuya existencia y propiedades no dependen del hecho de que sea tomada en consideración perceptiva y cognitivamente por ningún hablante. Por otro lado, el conocimiento obtenido perceptivamente es objetivo en la medida en que se trata de un conocimiento *impersonal*. Cualquier sujeto (cuyo sistema sensorial funcione en condiciones normales) habría tenido las mismas percepciones visuales que yo he tenido delante de la pared blanca y dichas percepciones habrían generado en él las mismas creencias que generan en mí. Sin embargo, no encontramos en el auto-conocimiento ninguno de estos rasgos vinculados a la objetividad, ya que, en primer lugar, no tiene sentido pensar en una contraposición entre nuestra mente y nuestros pensamientos, lo cual *impide la existencia de brecha causal alguna* entre, por un lado, el contenido de mis creencias de primer orden y, por otro, los juicios que formo al conocer estas creencias. En segundo lugar, *el auto-conocimiento es un conocimiento estrictamente personal*, por cuanto que lo conocido y la acción de conocerlo se encuentran dentro de los límites del punto de vista de primera persona. En definitiva, nadie puede conocer por mí y de la manera en que yo lo hago qué es lo que creo, deseo o temo.

Burge emplea el modelo basado en la inclusión para mostrar que el planteamiento simple del conflicto entre el externismo y el acceso privilegiado yerra en su objetivo. Este planteamiento denunciaba que, de suscribir las tesis externistas, tendríamos que admitir que las relaciones epistémicas con nuestros propios contenidos mentales se convertirían en relaciones indirectas y problemáticas. La razón es que para saber lo que pensamos tendríamos que conocer previamente aquellos condicionantes empíricos de nuestras creencias y deseos. Sin embargo, Burge responde que no se requiere nada ajeno a la acción misma de pensar sobre nuestros pensamientos para realizar esa acción y tener garantías, además, de que se realiza con éxito. Las justificaciones empíricas que se necesitan para lograr un conocimiento de los objetos externos son accesorias a la hora de efectuar con éxito los actos de auto-conocimiento básico, y esto también incluye las justificaciones empíricas sobre la naturaleza de los objetos externos que determinan los contenidos de nuestros pensamientos según el externismo.¹⁰¹ Con respecto a este punto, Burge sostiene que esos condicionantes

¹⁰¹ John Heil ofrece un argumento adicional al respecto. Afirmar que es necesario conocer las condiciones de identidad de un contenido mental para poder tener acceso cognitivo a ese contenido supondría vernos involucrados en un regreso al infinito. De esta manera, supongamos que se requiere el conocimiento de I, la condición de identidad del contenido C, para conocer que C. Ahora bien, la actividad de conocer que C es, asimismo, un estado mental con contenido, C', para cuyo conocimiento se requiere, *ex hypothesi*, un

empíricos (los géneros naturales presentes en el entorno, los usos lingüísticos de la comunidad) poseen un estatus de elementos *presupuestos* en los pensamientos de un hablante (Burge 1982, p. 91 y 1988a, 653). Para comprender esto, recordemos que la localización externa de los factores naturales o sociales que determinan los contenidos de nuestros pensamientos implica que un sujeto podría desconocer dichos factores. Más concretamente, este hablante podría ignorar la diferencia de naturaleza que existe entre un condicionante externo actual y un condicionante externo posible tal que la presencia de uno u otro en su entorno supondría una diferencia entre sus contenidos psicológicos actuales y posibles. La presuposición, por consiguiente, no tiene una naturaleza epistémica, sino que su carácter metafísico queda patente desde el momento en que el externismo establece que la *naturaleza o esencia* de las creencias de un hablante *depende constitutivamente* de la presencia de un condicionante natural o social concreto en su entorno. Esto es, al mantener que la creencia de que el agua carece de color presupone la existencia del agua, los externistas afirman que la existencia de ese líquido configura la naturaleza de la creencia en tanto que creencia acerca del agua (y no acerca de otro líquido distinto, como el que denominan “agua” los habitantes de la Tierra Gemela).

Aclarado esto, Burge se pregunta de qué manera el modelo que él mismo propone para explicar el auto-conocimiento podría hacer frente a una situación como la siguiente (Burge 1988a, pp. 652 y ss y 1988b, p. 213). Supongamos que Óscar₁, el habitante de la Tierra, es trasladado sin que él lo sepa a la Tierra Gemela. Atendiendo a las premisas del experimento mental de Putnam, este sujeto no percibiría ninguna diferencia entre su vida en la Tierra y su vida en el nuevo destino porque, por hipótesis, todo en la Tierra y en la Tierra Gemela le parece lo mismo. Sin embargo, nosotros sabemos que no es así, porque los líquidos que Óscar₁ y Óscar₂ denominan “agua” son líquidos esencialmente distintos, aunque indistinguibles a simple vista. Según lo defendido por las tesis externistas, el cambio de entorno y el consiguiente cambio en los

conocimiento previo de aquel factor, I', que lo determina. Sin embargo, si así fuera, estaríamos avanzando por una vía peligrosa, porque los contenidos generados en cada acto de auto-conocimiento se multiplican hasta el infinito. Heil se expresa de esta manera: “[c]ualquiera que lo pregunte [si el conocimiento de un contenido mental requiere el conocimiento de los elementos que lo determinan] debe enfrentarse al fantasma de un regreso: si el hecho de que mi pensamiento tenga un contenido particular requiere que yo reconozca que se dan ciertas condiciones, entonces esto requeriría que yo tuviese algún otro pensamiento con un contenido particular, es decir, con un contenido correspondiente a ese reconocimiento. Pero, por supuesto, este pensamiento podría requerir su propio reconocimiento correspondiente de que se dan las condiciones apropiadas para su contenido, y nos veríamos involucrados en un regreso” (Heil 1988, p. 250).

géneros naturales conllevarían una modificación en la identidad de los contenidos psicológicos de Óscar₁. Es decir, que si la identidad de los contenidos y conceptos de un hablante depende, en parte, del entorno natural o social en que está inserto, entonces un cambio de entorno – distinto en un sentido relevante – entraña un cambio de contenidos y conceptos. Burge nos advierte de que la variación en los contenidos de Óscar₁ no sería inmediata, aunque si este sujeto permaneciera en la Tierra Gemela el suficiente tiempo como para establecer relaciones estables con el líquido que los habitantes de ese entorno llaman “agua”, las creencias que Óscar₁ expresaría mediante la emisión de oraciones como “el agua es húmeda”, “la sequía consiste en escasez de agua”, etc., serían creencias que no contendrían el concepto de agua.

Ahora bien, tal y como está diseñado el experimento, parece que Óscar₁ no advertiría el cambio que se ha producido en el concepto que él expresa mediante el uso del término “agua”, dado que él no ha tenido noticia ni del traslado ni de la diferencia en la naturaleza de los líquidos relevantes. Por ello, si se llevaran a cabo una serie de traslados de la Tierra a la Tierra Gemela y viceversa, tales que en cada uno de ellos Óscar₁ permaneciera el tiempo suficiente como para adquirir, cada vez, un concepto distinto y después se informara a este hablante de que tales traslados han tenido lugar, es probable pensar que Óscar₁ sería incapaz de dar una respuesta satisfactoria a la pregunta “¿puedes identificar en qué momento ha tenido lugar un cambio en los conceptos que expresas mediante el término ‘agua’?”. La cuestión que merece la pena responder es si la incapacidad de Óscar₁ para fechar el cambio conceptual supone un atentado contra el acceso privilegiado a sus propios conceptos y contenidos mentales. Estos planteamientos incompatibilistas, que ponen de manifiesto la falta de habilidades discriminativas en sujetos que viajan inadvertidamente a otros entornos, han recibido el nombre de *casos de traslado lento* (“slow switching cases”). Aunque nos ocuparemos de estos casos con mayor detenimiento en el próximo capítulo conviene aludir a ellos en la presente sección, en la medida en que Burge considera que el modelo basado en la inclusión bloquea tanto el planteamiento incompatibilista simple como la versión sofisticada de este problema que son los casos de traslado. En este sentido, este autor admite que ni Óscar₁ ni ningún otro hablante sería capaz de discriminar por mera introspección si se encuentra en la situación actual o en una situación gemela donde todo le parecería igual aunque algunos de sus conceptos y creencias fueran distintos (Burge 1988a, p. 652 y 1988b, p. 213). Sin embargo, aunque carezcamos de la habilidad para discriminar entre los conceptos que expresaríamos en una situación u otra, para

Burge esta concesión no pone en peligro el acceso privilegiado. La inclusión conceptual implicada en el auto-conocimiento propicia que las condiciones de identidad (parcialmente externas) del contenido de nuestros juicios de primer orden sean también las condiciones de identidad del contenido de nuestros juicios de segundo orden. Si Óscar₁ expresa el concepto de agua al decir “el océano está lleno de agua”, entonces el conocimiento que este sujeto puede obtener de sus propias creencias vendrá establecido por juicios de segundo orden en los que figura, también, el concepto de agua. Aceptando el externismo, si este agente fuera trasladado a la Tierra Gemela y estableciera relaciones sólidas con ese nuevo entorno, al emitir una preferencia como “el océano está lleno de agua” la creencia expresada mediante esta oración no contendría el concepto de agua, sino un concepto distinto, que podemos denominar “concepto de agua₂”. Si reflexionara sobre dicha creencia, nuestro hablante formaría un juicio de segundo orden que contendría también el concepto de agua₂. Siguiendo las pautas dadas por Burge, Óscar₁ siempre sabe lo que cree en casos básicos como éste porque para averiguarlo le basta con enjuiciar reflexivamente sus creencias relevantes, distintas en cada entorno.

Comprobamos, así, que la diferencia entre los entornos no sólo afecta a las creencias de primer orden, haciendo que los pensamientos de Óscar₁ en cada entorno versen sobre líquidos distintos, sino que también afectará a los juicios de segundo orden mediante los cuales este hablante se adscribe sus propias creencias. Aunque el cambio conceptual le pase tan inadvertido a Óscar₁ como el traslado mismo, esto no impide que sus juicios de tipo *Cogito* sean siempre verdaderos y estén justificados *en el momento de realizarlos*, ya sea que dicha realización tenga lugar en la Tierra o en la Tierra Gemela (Burge 1988a, p. 659).¹⁰²

El modelo basado en la inclusión presentado por Burge para hacer frente al desafío incompatibilista simple ha recibido apoyo de autores externistas muy importantes.¹⁰³ No obstante, muchos partidarios del incompatibilismo no se dieron por vencidos con la respuesta de Burge y, de hecho, las deficiencias del modelo basado en la inclusión han servido de base para elaborar planteamientos incompatibilistas más complejos, como veremos en los próximos capítulos. De hecho, el punto en el que se

¹⁰² Burge sostiene que si alguien informara a Óscar₁ de que esos traslados han tenido lugar y este individuo se preguntara a sí mismo qué concepto está usando en ese preciso instante, Óscar₁ se respondería a sí mismo que ambos. La nueva información que le proporcionamos sobre el cambio conceptual que ha tenido lugar no le haría dudar sobre los conceptos que posee (Burge 1988a, p. 659).

¹⁰³ Cf. Davidson (1990), p. 271; Heil (1988), pp. 244 y ss.; Davies (2000), p. 391, Gibbons (1996), p. 290, Bernecker (1996), p. 338, Ludlow (1996), p. 312 y Shoemaker (1994), p. 260.

retomará la discusión tiene que ver con el siguiente problema: para que un sujeto tenga control epistémico privilegiado sobre sus contenidos y conceptos, ¿es necesario que este sujeto posea la capacidad de comparar y distinguir un concepto o un contenido de otro? Burge, según hemos visto, admitía que Óscar₁ era incapaz de discriminar su situación actual de una situación contrafáctica en la que, aunque todo le pareciera igual, sus conceptos y contenidos han cambiado. Sin embargo, para él esto no impide decir que Óscar₁ sabe lo que piensa en cada una de estas situaciones.

Una segunda idea de Burge que ha suscitado un gran debate tiene que ver con su afirmación de que nuestros estados mentales de carácter proposicional *presuponen metafísicamente* aquellas condiciones externas que determinan, en parte, su identidad. Muchos autores se preguntarán si dicha presuposición es de tal naturaleza que nos permite conocer la existencia de estas condiciones empíricas con la misma inmediatez, certeza y fiabilidad con la que conocemos nuestros pensamientos condicionados por ellas. La pregunta que se plantea y que examinaremos con detenimiento en el cuarto capítulo de este trabajo es la de si un partidario del externismo que, al mismo tiempo, respete la intuición del acceso privilegiado, se encuentra comprometido con la afirmación de que conocer de manera privilegiada nuestros pensamientos conlleva conocer de manera igualmente privilegiada los factores externos que nuestros pensamientos presuponen. Contentémonos, por el momento, con un planteamiento breve de estas dos cuestiones, de las que nos ocuparemos más adelante, y retenemos la idea fundamental sobre la que se sostiene el compatibilismo de Burge, a saber, que el auto-conocimiento es neutral con respecto a la individuación externista o no de nuestros contenidos.

3. *Donald Davidson: el externismo en el marco de la interpretación radical*

En esta sección examinaremos las ideas de Donald Davidson para compatibilizar el externismo psicológico con el conocimiento privilegiado de los propios pensamientos. El itinerario que seguiremos en esta tercera sección es el siguiente. Comenzaremos examinando la concepción del externismo sostenida por Davidson y qué relación tiene ésta con sus proyectos más importantes en el ámbito de la filosofía del lenguaje. Esta tarea resulta fundamental, debido a que su trabajo en semántica no coincide ni en sus presupuestos ni en sus conclusiones con las teorías externistas en semántica, teorías que, como se dijo, constituyen el zócalo de las concepciones más usuales del externismo

psicológico. El resto de esta sección estará dedicado a analizar el problema incompatibilista simple, tal y como es planteado por Davidson, concluyendo con la solución que propone el autor a este problema.

Dicho esto, expliquemos de qué manera entiende Davidson el externismo y qué papel juega esta teoría de la individuación en el marco de sus consideraciones filosóficas. El externismo constituye, para él, un requisito indispensable a la hora de emprender nuestras investigaciones en semántica y psicología, pues sólo si tomamos en consideración los factores externos que han *causado* las preferencias y pensamientos de un hablante podremos *identificar* el significado de sus palabras y el contenido de sus pensamientos. Desde la comprensión davidsoniana, el externismo es un presupuesto que sustenta la práctica totalidad de nuestra vida lingüística y psicológica. Así, no sólo hemos de contemplar la contribución de factores externos a la hora de explicar el comportamiento semántico de ciertos términos lingüísticos (nombres propios, indécicos y términos generales), sino que la individuación externista afecta al significado de nuestras preferencias oracionales de manera general (Davidson 1988a, p. 84). En el ámbito de las atribuciones psicológicas, este autor percibe también una dependencia global de nuestros conceptos y contenidos mentales respecto del entorno en el que han sido adquiridos. Por todo lo dicho, Davidson afirma que sólo seremos capaces de averiguar qué piensan y qué quieren decir los demás identificando el contenido de sus pensamientos y el significado de sus palabras a partir de las condiciones externas que han intervenido en su causación.

Nos interesa destacar la gran importancia que Davidson reserva al externismo en el terreno psicológico y lingüístico porque ello muestra, por un lado, su compromiso con esta tesis y, por otro, la diferencia que existe entre su comprensión del externismo y la comprensión más usual del mismo, que reserva el tratamiento externista para los conceptos expresados por términos de género natural (Putnam) o por términos generales (Burge). Las discrepancias de Davidson con el externismo psicológico más usual tienen raíces profundas y conciernen a los supuestos semánticos de ambas visiones del externismo. Como hemos señalado, Davidson no es un seguidor de las teorías externistas en semántica, tal y como las hemos caracterizado en el capítulo anterior, y rechaza en gran medida los postulados y las nociones de las que hacen uso estas teorías. Por ejemplo, él rehuye cualquier examen del significado de los términos lingüísticos aislados de su contexto oracional (Davidson 1965, p. 10) y cuestiona tanto el esencialismo que está presente en algunas versiones del externismo semántico como la

misma noción de designación rígida, un concepto central en algunas de estas versiones. Aunque una explicación detallada de las propuestas semánticas de Davidson rebasaría los límites de nuestro trabajo, conviene indicar algunas de sus ideas principales, especialmente en lo que se refiere a la cuestión de la interpretación radical.

Comencemos señalando que por *interpretación radical* Davidson entiende fundamentalmente el hecho de saber qué quieren decir los demás hablantes mediante sus emisiones lingüísticas, puesto que “[t]oda comprensión del habla de otro implica interpretación radical” (Davidson 1973, p. 374). Nuestros intercambios lingüísticos cotidianos pueden contemplarse bajo el signo de la interpretación, aunque usualmente este aspecto pase desapercibido porque en nuestras relaciones con los demás hablantes es habitual presuponer que compartimos el mismo lenguaje y, por tanto, que no hay nada que interpretar. El programa que Davidson desarrolla, inspirado en la traducción radical quiniana,¹⁰⁴ tiene como fin desvelar qué condiciones hacen posible la interpretación y, para lograrlo, este autor parte de una situación idealizada, en la que un hablante trata de comprender a otros hablantes cuyo lenguaje le resulta por completo desconocido y donde el cariz interpretativo de la comprensión lingüística resulta mucho más evidente. Respetando la costumbre al uso, denominaremos “nativos” a los hablantes cuyas preferencias se busca interpretar. La idea rectora de la interpretación radical queda establecida del siguiente modo, a saber, sólo mediante la elaboración de una teoría que nos proporcione el significado de las preferencias emitidas por los hablantes nativos podremos interpretar dichas preferencias, es decir, podremos averiguar qué significan las oraciones de ese lenguaje. Así, el sujeto que lleva a cabo la interpretación radical debe servirse de una teoría del significado que le permita obtener aquellos teoremas en los cuales se vinculan las preferencias de los hablantes nativos con sus correspondientes interpretaciones en el lenguaje del intérprete, siempre bajo el requisito de que los elementos así vinculados compartan sus condiciones de verdad.¹⁰⁵

¹⁰⁴ En este sentido, Davidson reconoce su profunda deuda con Quine, aunque señale con frecuencia las diferencias existentes entre su proyecto y el quiniano. Por ejemplo, Davidson indica que hablar de “interpretación” y no de “traducción” permite destacar los aspectos semánticos del lenguaje. Por otro lado, Davidson es mucho más liberal a la hora de autorizar la clase de evidencias de las que puede servirse un intérprete en su tarea, mientras que Quine sólo toma en consideración el comportamiento de los hablantes y las estimulaciones perceptivas proximales.

¹⁰⁵ Este aspecto de la empresa de interpretación radical puede clarificarse tomando en consideración las bases de la teoría del significado davidsoniana para lenguajes naturales. En primer lugar, para este autor cualquier teoría aceptable a este respecto debe atender a dos características de cualquier lenguaje natural como, por ejemplo, L. Por un lado, los lenguajes naturales han de poder ser aprendidos por seres finitos, es decir, seres cuyas capacidades intelectuales y cuyo tiempo de vida son limitados. Por otro lado, un lenguaje natural contiene un número potencialmente infinito de oraciones con significado. Estas dos

Para obtener estos teoremas, el intérprete necesita observar el comportamiento de los nativos y relacionar, así, las preferencias que estos emiten con cambios en las circunstancias *externas* observables. Asimismo, cabe el recurso quiniano al asentimiento provocado, es decir, que el intérprete puede preguntar al hablante sobre el significado de su preferencia al emitirla mientras alude ostensivamente a la situación que, según la hipótesis del intérprete, constituye su significado.

Uno de los pivotes de la interpretación radical consiste en la vinculación que se presume entre las aserciones y las creencias de los hablantes. Presumimos que si un nativo asiente a una preferencia y, por tanto, la toma por verdadera, es legítimo atribuirle una actitud cognitiva favorable hacia el contenido proposicional expresado por la oración. Por otro lado, debemos mencionar que un requisito metodológico fundamental para llevar a cabo con éxito la interpretación radical consiste en el respeto al Principio de Caridad. Aunque hemos aludido a dicho Principio en la introducción de este trabajo, resulta conveniente esclarecer cuáles son para Davidson las concreciones del mismo y mostrar, con ello, la relación existente entre el proyecto de interpretación

peculiaridades de los lenguajes naturales imponen dos requisitos a toda teoría que quiera dar cuenta del significado de sus oraciones, a saber, que dicha teoría debe poner en juego recursos limitados (recursos que puedan manejar los seres humanos) y, además, que debe ser lo suficientemente potente como para proporcionar el significado de las infinitas oraciones que podrían formularse en el seno de un lenguaje natural concreto (Davidson 1967, p. 17). La idea de este autor es que estos dos requisitos sólo pueden aunarse en una teoría *composicional* del significado, diseñada de tal manera que nos permita dar cuenta del significado de las expresiones complejas de L (potencialmente infinitas) a partir del significado de las expresiones primitivas (y en número finito) de L. Davidson analiza, además, la forma que han de tener las especificaciones del significado de las oraciones de L que debería proporcionarnos nuestra teoría. Por un lado, deben vincular las oraciones de L con aquellas oraciones que nos proporcionen su significado. Por otro lado, la vinculación entre una oración de L y la oración (formulada en el lenguaje de la teoría) que nos da su significado tiene que ser una vinculación establecida mediante una conectiva apropiada y en la que figure una noción extensional en lugar de “significa que”. De acuerdo con dichas pautas, la forma de las especificaciones del significado de las oraciones de L que esperamos que nos proporcione esa teoría es la siguiente: (S) *O* es *V* si y solo si *P*. Davidson denomina “teoremas” a estas especificaciones cuya forma acaba de ser indicada. (S) es una oración del lenguaje en el que está formulada nuestra teoría y que puede ser distinto o no del lenguaje L. En cualquier caso, debemos distinguir dos niveles de lenguaje. En primer lugar, el lenguaje en el que formulamos nuestra teoría del significado es un *metalenguaje*, esto es, un lenguaje que utilizamos para hablar de otro lenguaje, que, en nuestro caso, es L. En segundo lugar, tomamos a L, el significado de cuyas oraciones queremos especificar, como *lenguaje objeto*, es decir, como un lenguaje del que hablamos haciendo uso de los recursos del metalenguaje. En (S), “O” ocupa el lugar del nombre metalingüístico de una oración del lenguaje objeto, mientras que “P” representa la oración metalingüística que especifica el significado de “O”. “V” es, por su parte, la noción que reemplazará a “significa que”. La pregunta crucial es la siguiente: ¿qué propiedad designa “V”? Debemos tener en cuenta que dicha propiedad, cualquiera que sea, ha de permitir que “O” sea equivalente a “P” en cuanto a su valor de verdad, puesto que la conectiva que vincula estas oraciones es un bicondicional. Davidson califica su propia respuesta de “simple y radical”: “V” representa el predicado “es verdadera”. De esta forma, una teoría del significado para un lenguaje natural L es una teoría que debe implicar, para todas las oraciones de ese lenguaje, teoremas de la forma (S) en los que se especifica el significado de cada una de estas oraciones indicando las condiciones que deberían darse para que estas oraciones fueran verdaderas.

radical y la comprensión davidsoniana del externismo. En este sentido, el Principio de Caridad se desglosa en dos Principios adicionales, el *Principio de Correspondencia* y el *Principio de Coherencia* (Davidson 1991, p. 288). En primer lugar, por lo que respecta al Principio de Correspondencia, hemos de suponer que en los casos más básicos nuestras preferencias y pensamientos se refieren a aquellos “rasgos del mundo” con los que están causalmente conectados.¹⁰⁶ El intérprete puede servirse de esta correspondencia para establecer una correlación entre las preferencias y creencias del nativo con acontecimientos y estados de cosas objetivos (observables por él mismo y por el nativo), bajo la hipótesis de que constituyen el contenido de las creencias que los hablantes expresan mediante sus preferencias. El externismo, entendido como *identificación de contenidos* (semánticos y psicológicos) a partir de elementos externos al hablante, se erige, como vemos, en un requisito ineludible de la interpretación radical. En lo referente al Principio de Coherencia, un intérprete debe suponer que los nativos están dotados de sistemas de creencias lógicamente coherentes y globalmente verdaderos. Sería imposible concluir con éxito una interpretación radical o bien bajo la hipótesis de que las oraciones de los nativos son por lo general falsas, o bien suponiendo que estos hablantes mantienen multitud de creencias contradictorias.

En relación con esto, el *factum* de la asimetría entre la primera y la tercera persona también interviene en calidad de presupuesto de la interpretación radical, hasta el punto de que Davidson explica la autoridad fundamental del auto-conocimiento en las coordenadas de la interpretación. Así, considera que cuando interpretamos a un hablante cuyo lenguaje desconocemos no tiene sentido cuestionar si este hablante realmente sabe lo que piensa o lo que significan sus palabras. Por el contrario, tomar a un sujeto como un hablante racional implica *per se* reconocer en él tanto la capacidad de dominar su

¹⁰⁶ En un caso extremo, un individuo que no haya entablado contactos causales con elementos externos sería un individuo carente de vida cognitiva y cuyas preferencias carecerían de contenido semántico. Davidson propone al respecto el experimento mental del “hombre del pantano” (Davidson 1987 p. 46), una réplica interna exacta de sí mismo, surgida espontáneamente por el efecto de un rayo. Por lo que sabemos, la historia causal del hombre del pantano es nula. Él nunca conoció a Quine, ni participó en la Segunda Guerra Mundial, ni impartió clases en la Universidad de California, cosas que Davidson sí hizo. Lo que este autor pretende demostrar es que dos individuos intrínsecamente idénticos pero cuyos contactos causales con elementos externos hayan sido diferentes ameritan distintas descripciones de su vida cognitiva. Así, mientras que Davidson sí cree que Quine es un gran filósofo, no podemos atribuir la misma creencia al hombre del pantano.

propio lenguaje como la posesión de una autoridad especial a la hora de realizar auto-atribuciones.¹⁰⁷

Una vez que hemos recogido lo más esencial del proyecto davidsoniano de la interpretación radical, podemos extraer dos conclusiones importantes para el problema que nos ocupa. La primera conclusión es que el punto de vista externista adopta un gran protagonismo en las consideraciones de Davidson, al ser uno de los requisitos indispensables para realizar exitosamente la interpretación radical. De esta manera, sólo podremos saber qué dicen y piensan los hablantes a partir de la *triangularidad mente-lenguaje-mundo*, formada por la identificación sistemática entre, por un lado, los acontecimientos y estados de cosas objetivos y, por otro, el contenido semántico y psicológico causado por dichos rasgos objetivos.¹⁰⁸ La segunda conclusión tiene que ver con el hecho de que la asimetría entre la primera y la tercera persona también figura como una de las condiciones de posibilidad de la interpretación radical. A partir de esto, podemos conjeturar que si existiera un conflicto entre el externismo y el acceso privilegiado a los propios contenidos mentales, la viabilidad del programa filosófico de Davidson se vería seriamente debilitada, dada la centralidad que estas tesis exhiben en dicho programa.

En lo que sigue, examinaremos la vía davidsoniana para hacer frente al planteamiento incompatibilista simple, para lo cual conviene establecer la pauta que sigue dicho planteamiento en la obra de este autor (Davidson 1987, p. 62):

(Dav-1) El externismo pretende “sacar” nuestros pensamientos de nuestras cabezas.

(Dav-2) Saber lo que pensamos requiere captar nuestros pensamientos.

(Dav-3) Si “sacamos” nuestros pensamientos de nuestras cabezas, entonces no podremos captarlos.

(Dav-4) Por tanto, el externismo atenta contra la condición del auto-conocimiento.

La estrategia de Davidson para enfrentarse a este argumento se basa en el cuestionamiento de las premisas (Dav-1) y (Dav-2). En particular, sus esfuerzos se dirigen a mostrar, contra (Dav-1), que su versión del externismo no implica “sacar” los pensamientos de nuestras cabezas de una manera que ponga en peligro el acceso

¹⁰⁷ Algunos autores discrepan de Davidson en este aspecto al considerar que *presuponer* la asimetría como condición de posibilidad de la interpretación radical no es *explicar* en absoluto el hecho de la asimetría. Cf. (Lepore/Ludwig 2005, p. 368)

¹⁰⁸ Cf. Davidson (1990), pp. 277 y ss. y (1991), pp. 280-281.

privilegiado y cuestionar, contra (Dav-2), que el auto-conocimiento no tiene lugar por medio de captación alguna.

Centrémonos, por el momento, en las alegaciones de Davidson contra la primera premisa del argumento incompatibilista. A partir de ellas este autor pretende hacer plausible que su propia comprensión del externismo, al contrario que la de Burge, no implica “situar fuera de la cabeza” los pensamientos de forma que un sujeto sea incapaz de saber lo que piensa. Para apreciar mejor el contraste entre ambas formas de entender el externismo, atendamos primero a las críticas que Davidson dirige a las propuestas de Burge. Según el primer autor, el externismo de Burge atenta contra el acceso privilegiado a los propios pensamientos en cuanto que legitima atribuciones conceptuales a sujetos que sólo dominan en parte los conceptos que les atribuimos. Recordemos que en el experimento mental en torno a la noción de artritis que Burge planteó en (1979), Bert, el hablante de nuestra comunidad lingüística, tenía creencias, deseos y temores en torno a la artritis a pesar de malentender parcialmente este concepto. El problema consiste, según Davidson, en que si un hablante puede usar términos lingüísticos sin saber, en parte, qué quieren decir, y si, además, puede tener nociones que sólo otros dominan correctamente, de esto se desprende que el significado de sus palabras y el contenido de sus pensamientos escapan a su control epistémico. Por este motivo, Davidson considera que debemos rechazar la viabilidad del externismo en la versión propuesta por Burge.¹⁰⁹

Por otro lado, Davidson rechaza cierta interpretación de las conclusiones que Putnam obtiene en su artículo de (1975c), interpretación que, siguiendo a Bilgrami, denominaremos “visión bifurcada del contenido” (Bilgrami 1992a, pp. 3 y ss.). Esta

¹⁰⁹ Davidson propone un ejemplo para mostrar de qué manera las atribuciones autorizadas por el externismo de Burge atentan contra el auto-conocimiento. En el ejemplo de Davidson, un individuo llamado Smith informa a un tercero de que Arthur y el propio Davidson creen que Carl tiene artritis. Por hipótesis, el conocimiento que tienen Arthur y Davidson de esta enfermedad es distinto. Mientras que Davidson tiene la creencia errónea de que el término “artritis” sólo se aplica a las inflamaciones de las articulaciones causadas por depósitos de calcio, Arthur cree (con verdad) que ese término se aplica a cualquier inflamación de las articulaciones. ¿Sería correcto, pregunta Davidson, decir “Davidson y Arthur creen que Carl tiene artritis”? Según este autor, este informe no resultaría “del todo correcto” (Davidson 1987, p. 58). Por tanto, sin matizaciones adicionales, no deberíamos atribuir tanto a Davidson como a Arthur creencias sobre la artritis, visto que su dominio de la noción relevante difiere. No obstante, Owens ha señalado que un externista tipo Burge admitiría que el informe “Davidson y Arthur creen que Carl tiene artritis” no es, en efecto, del todo correcto, siempre y cuando esto no implique rechazar la verdad del informe. Está claro que en algún sentido Davidson y Arthur no quieren decir lo mismo (“mean”) cuando profieren la oración “Carl tiene artritis”, visto su distinto dominio de la noción de artritis. Sin embargo, atribuir a ambos pensamientos sobre la artritis no es una práctica que pase totalmente por alto sus intenciones y su competencia. En este sentido, es cierto que Davidson cree que Carl tiene artritis y para decirlo resulta suficiente que Davidson tenga un dominio básico (aunque incompleto) de la noción relevante. Cf. Owens (2003), p. 209.

visión toma pie en la distinción de Putnam entre estados psicológicos en sentido restringido y estados psicológicos en sentido amplio y propone, de acuerdo con esto, la existencia de dos tipos de contenido correspondientes. Mientras que el contenido en sentido restringido está determinado exclusivamente por factores que los sujetos conocen plenamente por ser internos a su cabeza, el contenido amplio recibe una individuación externista que a menudo supone eliminar su estatus de objeto de conocimiento privilegiado por parte de los hablantes que los poseen. La interpretación bifurcacionista del experimento mental de la Tierra Gemela atribuye a Óscar₁ y Óscar₂ idénticos contenidos en sentido restringido, puesto que sus creencias son indistinguibles desde el punto de vista interno a cada uno de ellos, pero distintos contenidos en sentido amplio, dada la diferencia existente en sus respectivos entornos. No obstante, para Davidson, la visión bifurcada del contenido tiene su origen en una concepción errónea de la mente, precisamente aquella que entiende el auto-conocimiento bajo el signo de la “captación” – como reza la premisa (Dav-2) del argumento incompatibilista – y que, como veremos, será objeto de un fuerte cuestionamiento por parte de Davison.

La interpretación que propone este autor del experimento mental de Putnam es diferente. Por un lado, dice Davidson, nuestras prácticas de atribución nos compelen a emplear distintas *descripciones* para dar cuenta de los pensamientos de Óscar₁ y Óscar₂. En la descripción utilizada para informar de la creencia de Óscar₁ figurará el término “agua”, puesto que ese líquido es el rasgo objetivo que pertenece a la historia causal de este hablante. Como la historia causal de Óscar₂ es diferente en ese sentido, no podremos emplear la misma descripción para informar acerca de sus creencias. Ahora bien, dado que Óscar₁ y Óscar₂ son hablantes internamente idénticos, sus estados mentales, en tanto que estados internos físicos, también lo serán. Es decir, que la diferencia que existe en sus respectivos entornos no conlleva una diferencia en sus estados mentales *considerados como eventos neuronales*.

En relación a esto, la distinción de Owens entre dos comprensiones distintas del externismo nos permite profundizar en la interpretación davidsoniana del experimento mental de Putnam (Owens 2003, pp. 204-205). Por un lado, Owens ve en Davidson una defensa del *externismo lingüístico*, la tesis que establece que las diferencias relevantes en el entorno externo a los sujetos conllevan diferencias en las descripciones que empleamos para caracterizar sus estados psicológicos. Por otro lado, el *externismo*

metafísico, que Davidson rechaza, va un tanto más lejos y afirma que esas diferencias en el entorno extra-mental suponen una diferencia en los estados psicológicos mismos.¹¹⁰

Como vemos, el intento de Davidson para impedir el establecimiento de la primera premisa del argumento incompatibilista – que el externismo pretende “sacar” nuestros pensamientos de nuestras cabezas – tiene que ver con su apoyo a la comprensión lingüística del externismo psicológico y el consiguiente rechazo a la variante metafísica de esta tesis. La creencia de Óscar₁ en la humedad del agua está “dentro de su cabeza” porque es idéntica a uno de sus eventos neuronales. Y el hecho de que podamos usar distintas descripciones para capturar dicha creencia, atendiendo a la historia causal del hablante en cuestión, no altera en absoluto la identidad de los estados psicológicos de Óscar₁ y Óscar₂ y su identificación con un mismo movimiento neuronal, perfectamente interno a sus cabezas. Davidson concluye que su estrategia le permite eliminar “un obstáculo no basado en evidencias de nuestras actitudes proposicionales usuales, ya que si nuestras creencias usuales y otras actitudes pueden estar en la cabeza, entonces la amenaza a la autoridad de primera persona no puede provenir simplemente del hecho de que los factores externos sean relevantes para la identificación de las actitudes” (Davidson 1987, p. 66).

En nuestra opinión, el cuestionamiento davidsoniano de la primera premisa del argumento compatibilista está lejos de ser satisfactorio. Tenemos la impresión de que Davidson malinterpreta el sentido de la objeción en tanto que la naturaleza de la misma es epistémica y la réplica que ofrece este autor consiste en un examen de la identidad metafísica de los estados mentales. En este sentido, Davidson rechaza que la identificación externista de nuestros pensamientos suponga desalojar a éstos de nuestras

¹¹⁰ Para aclarar esta distinción tanto como sea posible, podemos examinar un ejemplo que nos ofrece el propio Davidson. En él compara la (distinta) identificación de dos estados mentales atendiendo a su distinto origen causal con la (distinta) identificación de dos quemaduras en la piel provocadas por dos factores externos diferentes (Davidson 1987, p. 67 y 1988a, p. 85). Imaginemos dos individuos, Juan₁ y Juan₂, idénticos molécula a molécula, que poseen, cada uno de ellos, una quemadura en su brazo derecho. Mientras que la quemadura de Juan₁ ha sido provocada por el sol, la quemadura de Juan₂ ha tenido una causa distinta. La idea de Davidson consiste en que la situación de Óscar₁ y de Óscar₂, los protagonistas de los experimentos mentales de Putnam, es similar a la situación de Juan₁ y Juan₂. Al igual que los pensamientos de Óscar₁ reciben una identificación distinta de la empleada para capturar los pensamientos de Óscar₂ sin que ello suponga una diferencia intrínseca entre ambos individuos, de la misma forma las afecciones en la piel de Juan₁ y Juan₂ son descritas como afecciones distintas sin que esa diferencia altere la identidad intrínseca exacta entre estos dos sujetos. La diferencia en las descripciones que utilizamos para capturar los pensamientos de Óscar₁ y Óscar₂ y para hablar de las quemaduras de Juan₁ y Juan₂ radica en la diferencia existente en los factores externos que han causado tales pensamientos y tales afecciones. Ahora bien, en tanto que la distinta identificación de las quemaduras de Juan₁ y Juan₂ a partir de diferentes factores externos no conlleva situar fuera de la piel dichas quemaduras, la diferente identificación de los pensamientos de Óscar₁ y Óscar₂ a partir de los distintos elementos externos que han intervenido en su causación no supone tampoco situar sus pensamientos fuera de su cabeza.

cabezas de modo tal que estos pensamientos dejen de ser eventos internos, idénticos a eventos neuronales. La creencia de Óscar₁ en la humedad del agua es “interna” a su cabeza, aunque la caractericemos en términos de factores externos a ella. Bajo esta lectura, los pensamientos de Óscar₁ y Óscar₂ son idénticos porque, por hipótesis, son eventos de sus cerebros, idénticos, por hipótesis. Ahora bien, no está claro que el proponente de la objeción incompatibilista simple pueda darse por satisfecho con semejante respuesta. Su preocupación tiene que ver con el hecho de que los protagonistas de los experimentos ignoran aquello en lo que consiste la diferencia entre sus entornos, ya que Óscar₁ y Óscar₂, recordemos, viven en mundos subjetivamente equivalentes. El “afuera de las cabezas” que pone en juego el experimento mental de la Tierra Gemela consiste, precisamente, en que aquello que lleva a establecer una diferencia en los estados mentales de sus protagonistas (o en las descripciones de esos estados mentales) puede residir fuera del alcance epistémico de los hablantes. Esto nos permite formular la objeción incompatibilista de una manera más precisa. Si los contenidos mentales de un sujeto pueden depender, en parte, de aspectos externos que trascienden su conocimiento, ¿puede seguir manteniendo este sujeto un control epistémico sobre dichos contenidos mentales? La réplica de Davidson no es una respuesta a dicha pregunta, sino a una muy diferente: el eslogan externista “los pensamientos no están en nuestras cabezas”, ¿conlleva que los pensamientos pueden no ser idénticos a eventos que suceden dentro de nuestras cabezas? Al confundir estas preguntas, el argumento de Davidson para reservar una localización neuronal y, en este sentido, “interna” para los pensamientos, no implica situar a éstos en un emplazamiento a partir del cual los sujetos puedan obtener réditos epistémicos. Por todo esto, que Davidson alegue que un pensamiento es (metafísicamente) idéntico a un estado cerebral resulta irrelevante para un debate que se centra en las consecuencias epistemológicas de la identificación de los contenidos mentales a partir de elementos que pueden ser desconocidos para el hablante a quien atribuimos dichos contenidos.¹¹¹

Ahora bien, a pesar de nuestras críticas es necesario reconocer que la réplica de Davidson a la premisa (Dav-2) del argumento incompatibilista simple es más sugerente y está mejor encaminada que su discusión de la premisa (Dav-1). Dicha réplica reza que el externismo psicológico no entra en conflicto con el auto-conocimiento entendido

¹¹¹ McKinsey hace una observación similar sobre las tesis de Davidson. Según este autor, la internalidad que Davidson reserva a los estados mentales no proporciona ningún argumento para reforzar la idea de que un hablante puede conocer de manera privilegiada que el episodio interno “es un episodio del pensamiento en vez de ser, por ejemplo, un episodio de indigestión” (McKinsey 1991, p. 11).

como “captación” de los propios pensamientos por la sencilla razón de que nunca ha existido “captación” alguna. Este autor participa, con esto, en las críticas a la visión perceptiva de la introspección que hemos recogido en la primera sección del presente capítulo. Según Davidson, toda teoría que intente explicar nuestra capacidad de autoconocimiento en términos de “captaciones” de “objetos” mentales es una teoría que se apoya en una imagen de la mente simplista y errónea, a saber, la imagen de “la mente como teatro” (Davidson 1987, p. 67 y ss.). En esta imagen “el ojo de la mente” ocupa el lugar del espectador, mientras que el escenario está integrado por nuestros pensamientos, ideas, datos sensoriales o impresiones, según sea una u otra la epistemología que dé apoyo a la teoría. Para este autor, la segunda premisa del argumento incompatibilista simple y la imagen de la mente en que se sustenta constituyen, en realidad, la fuente de todas las acusaciones de incompatibilidad entre el externismo y el acceso privilegiado. Se presume, desde dicha imagen, que saber lo que creemos consiste en capturar o contemplar un determinado objeto (nuestra creencia) cuya situación interna a nosotros mismos es la condición de posibilidad de su perfecto conocimiento. Por tanto, si el externismo hace depender el contenido mental de factores externos a la mente, el conocimiento de los contenidos se ve perturbado por la intervención de factores ajenos a ese “teatro” interno. Sin embargo, como dice Davidson de manera lacónica, “tener una creencia no es como tener un gato favorito” (Davidson 1997, p. 117). Esta afirmación se dirige contra todo intento de reificar aquello que es conocido por nosotros cuando sabemos qué deseamos, creemos o tememos, dado que sólo en base a tal reificación tiene sentido presumir la incompatibilidad entre el externismo y el acceso privilegiado. Davidson nos invita a rechazar el modo perceptivo de entender la introspección no sólo porque entra en conflicto con el externismo, sino porque este modelo genera por sí mismo graves dificultades. En este sentido, podemos apreciar que la imagen de la mente como teatro no puede ocultar su carácter metafórico y resulta muy problemático intentar precisar los términos de su formulación.¹¹²

Dicho esto, es preceptivo indicar cuáles son las propuestas positivas de Davidson en cuanto al modo efectivo en que tiene lugar el conocimiento de los estados mentales, propios y ajenos. Lo más llamativo, sin duda, es que dichas propuestas son

¹¹² Davidson realiza unas afirmaciones a este respecto con ciertos tintes de ironía: “los *voyeurs* sólo quieren tener representaciones ante el ojo de la mente, mientras que los más agresivos las asen; un inglés puede ser que meramente esté en contacto con los contenidos de la mente, mientras que los caracteres más afables los abrigan” (Davidson 1987, pp. 67-68). Además, la situación del yo en el modelo de la mente como teatro es sumamente ambigua, porque parece ocupar, simultáneamente, el papel del espectador, el rol de lo puesto en escena e incluso el lugar del propio teatro.

relativamente escasas y que, además, recogen lo esencial del “modelo basado en la inclusión” propuesto por Burge, a pesar de que Davison no mencione explícitamente su deuda con este autor.

Comencemos recordando que Davidson explica la asimetría entre el conocimiento de primera persona y el conocimiento de tercera persona en el marco de la interpretación radical. Sólo es posible interpretar a un hablante si presumimos en él la capacidad de saber lo que piensa y de saber lo que quiere decir (“mean”) mediante el uso de sus términos lingüísticos.¹¹³ El acceso privilegiado a los propios contenidos mentales y al significado de nuestras palabras es una condición necesaria para efectuar con éxito nuestro trato lingüístico con los demás hablantes, puesto que, recordemos, toda comprensión del lenguaje es interpretación radical. Pero no sólo eso. En alguno de sus textos, como (1997, p. 124), Davidson explica la asimetría entre las atribuciones de primera y tercera persona afirmando que en nuestras auto-adcripciones no existe un proceso de interpretación propiamente dicho, mientras que este proceso es inevitable en las atribuciones mentales a terceros. El resultado de esta diferencia es que los rasgos que caracterizan las tareas interpretativas emprendidas para averiguar qué piensan los demás no se presentan en el auto-conocimiento. Estos rasgos son el carácter *falible*, *tentativo*, *indeterminado* y *basado en evidencias* de las interpretaciones propuestas. Es necesario aclarar que la indeterminación de la interpretación aludida tiene un significado muy concreto.¹¹⁴ Supone que una misma actitud psicológica puede recibir distintas interpretaciones igualmente legítimas. Esto es así porque “nuestras oraciones proporcionan la única medida de lo mental” (Davidson 1997, p. 120). Y, al ser posible que más de una descripción resulte empíricamente adecuada con respecto a las evidencias que tomamos en cuenta para interpretar a un hablante, todas ellas puedan ser usadas para informar de sus pensamientos.¹¹⁵ Ahora bien, admitir la indeterminación de la interpretación no supone menoscabar, según Davidson, la entidad del contenido,

¹¹³ Ya vimos que el Principio de Caridad en su versión de Principio de Coherencia impone no sólo la presunción de que un individuo no está en general equivocado sobre sus propias creencias, sino que las creencias de dicho individuo, al menos en sus casos más básicos, son globalmente verdaderas. (Davidson 1990, p. 270 y 1986, pp. 195 y ss.)

¹¹⁴ Ese sentido tiene un claro parentesco con la tesis quineana de la “indeterminación de la traducción”. Esta tesis afirma que un hablante que intente llevar a cabo una traducción radical de un lenguaje que le resulta por entero desconocido usando únicamente los indicios que proporciona el comportamiento de los hablantes nativos, podría elaborar distintos manuales de traducción incompatibles entre sí, siendo cada uno de ellos perfectamente adecuado a las evidencias conductuales disponibles. Cf. Quine (1960), pp. 26 y ss.

¹¹⁵ Algo que no ocurre, según Davidson, en las auto-adcripciones, puesto que “una persona est[á] restringida a una única manera de interpretarse a sí misma” (Davidson 1997, p. 124).

reduciéndolo a un mero resultado de nuestras prácticas de interpretación. Para demostrarlo nos presenta una analogía interesante. La relación “más pesado que” puede ser medida bien en libras, bien en kilogramos y esto no implica (i) que debamos reconocer entidad a las libras, a los kilogramos ni al resto de las unidades de medida ni (ii) que debamos negársela a “lo” medido. De manera equivalente, es posible permitir que los contenidos mentales puedan recibir distintas interpretaciones sin sostener una posición antirrealista en psicología (Davidson 1991, pp. 292-293 y 1997, p. 125).

Por último, ya hemos mencionado que Davidson se adhiere al modelo basado en la inclusión elaborado por Burge en (1988), aunque no hace mención explícita de esta inspiración. Así, en Davidson encontramos afirmaciones de este tipo:

“[U]n intérprete puede descubrir, o suponer correctamente sobre la base de evidencias indirectas, cuáles son los factores externos que determinan el contenido del pensamiento de otra persona; pero puesto que estos factores determinan tanto los contenidos de los pensamientos como los contenidos de los pensamientos que uno cree que tiene (al ser éstos el mismo pensamiento), no hay lugar para un error acerca de los contenidos de nuestros propios pensamientos del tipo que puede darse con respecto a los pensamientos de otros” (Davidson 1990, p. 271).¹¹⁶

La adhesión de Davidson al modelo basado en la inclusión propuesto por Burge blindará frente al error nuestros pensamientos reflexivos. Incluso en los casos de traslado, el hablante siempre conocería lo que cree en el momento de conocerlo, a pesar de no advertir el cambio sufrido en los factores que causan sus creencias.¹¹⁷

Antes de comenzar la cuarta y última sección de este segundo capítulo, querríamos ofrecer un balance general en torno a la eficacia de las propuestas de Davidson para conjugar el externismo con el acceso privilegiado a los propios contenidos mentales. Como se dijo, la clase de localización interna que este autor reserva a nuestros estados mentales no constituye, en nuestra opinión, una localización

¹¹⁶ Otra evidencia textual que demuestra que Davidson se adhiere al modelo basado en la inclusión figura en un trabajo de Boghossian que cita un artículo de Davidson en estado de borrador. La cita es la siguiente: “[m]ostrar que no existe ningún conflicto [entre el externismo y el conocimiento del contenido] es básicamente simple. Depende de darnos cuenta de que lo que quiera que sea responsable de los contenidos de nuestros pensamientos (tanto si es conocido por nosotros como si no) es también responsable del contenido del pensamiento acerca del hecho de que tenemos el pensamiento” (Boghossian 1994, p. 35). Ahora bien, la redacción definitiva del artículo de Davidson recoge lo siguiente: “[l]a razón básica es que lo que determina los contenidos de los pensamientos también determina lo que el sujeto piensa acerca de esos contenidos” (Davidson 1988, p. 664).

¹¹⁷ Davidson trata los casos de traslado en (1989), pp. 102 y ss.

que sirva para desactivar las preocupaciones incompatibilistas. Esto es, quienes acusan al externismo de “sacar” nuestros pensamientos de nuestras cabezas, no considerarán convincente una respuesta como la de Davidson, que se contenta con alegar que, en la medida en que nuestros pensamientos son idénticos a ciertos eventos neuronales, tales pensamientos estarán “dentro” de nuestras cabezas. Hemos procurado explicar que la dicotomía interno-externo que atraviesa el debate internismo-externismo tiene su fundamento en consideraciones epistemológicas, como sin duda lo tiene también el debate entre los partidarios y detractores de la compatibilidad entre el externismo y el acceso privilegiado. De ser como decimos, Davidson yerra en el tiro al argumentar a favor del compatibilismo, puesto que un sujeto no gana nada en términos de auto-conocimiento al saber que sus pensamientos, por ser idénticos a determinados movimientos neuronales, no están en otra parte que en su cabeza.

4. *El trilema de Bilgrami: externismo ortodoxo y atribuciones inconsistentes*

Las reflexiones de Bilgrami tienen un indudable interés para todo aquél que se ocupe del debate sobre la compatibilidad entre el externismo y el acceso privilegiado. Desde una posición que el autor califica de externista, pretende elaborar un modelo de atribución de contenidos mentales que no adolezca de los mismos defectos imputables al “externismo ortodoxo”. Este autor intentará demostrar que dicha concepción ortodoxa, que no es más que el externismo psicológico apoyado en las ideas semánticas de Putnam y las tesis anti-individualistas de Burge, legitiman atribuciones conceptuales inconsistentes, es decir, atribuciones que violentan nuestra consideración de los sujetos que reciben las atribuciones como agentes racionales. El impacto de este atentado contra el Principio de Caridad sólo puede reducirse o bien al precio de despojar al hablante de la capacidad de conocer sus propios contenidos, o bien suscribiendo una visión bifurcada del contenido mental. A lo largo de esta sección desplegaremos el argumento de Bilgrami para obtener las tres alternativas de su desafío disyuntivo, valoraremos la capacidad de los externismos ortodoxos para encontrar una respuesta al mismo y, por último, examinaremos la viabilidad de su propia comprensión del externismo.

Comenzaremos examinando el argumento de Bilgrami para revelar la incompatibilidad entre el externismo ortodoxo y el acceso privilegiado a los propios contenidos mentales. La primera premisa del trilema reza que esta comprensión del externismo da lugar a atribuciones inconsistentes. La justificación que ofrece Bilgrami

para establecer esta premisa tiene que ver con un rasgo que, a su juicio, se encuentra en la base de los externismos ortodoxos, a saber, su compromiso con lo que él denomina “la teoría de la referencia directa”. Debe decirse, sin embargo, que la caracterización que este autor ofrece de esta teoría no coincide con la que propusimos al respecto en el primer capítulo de este trabajo y más adelante tendremos oportunidad de aclarar algunas de las diferencias más importantes. Por este motivo y en aras de la claridad, la expresión “teoría de la referencia directa” cuando sea usada en el sentido de Bilgrami figurará entrecomillada y en minúsculas.

De acuerdo con Bilgrami, esta teoría sostiene que el significado de nuestros términos y los conceptos que expresamos mediante el uso de los mismos vienen especificados en función de las *naturalezas* o *esencias objetivas* que constituyen su extensión. Qué significado tiene el término “T” usado por el hablante S y qué concepto expresa S mediante el uso de “T” dependerá, según esta teoría, de las relaciones causales establecidas entre S y la entidad a la que “T” se refiere, si “T” es un término singular, o las entidades que forman parte de la extensión de “T”, si es un término general. En este último caso, un objeto o individuo pertenecerá a la extensión de “T” en virtud de compartir una misma esencia o naturaleza subyacente. Según Bilgrami, la “teoría de la referencia directa” considera que estas relaciones causales constituyen la condición necesaria y suficiente para la determinación de los significados y conceptos. Si bien puede ocurrir que ciertas creencias sean necesarias para fijar la referencia de “T”, no se requiere mediación conceptual alguna para usar posteriormente el término “T” con éxito y referirnos, así, a su portador, si “T” es un término singular, o a las entidades que forman parte de su extensión, si “T” es un término general. Es más, para esta “teoría de la referencia directa” las representaciones cognitivas que los hablantes asocian con el concepto expresado por “T” no juegan ningún papel a la hora de individuar dicho concepto (Bilgrami 1992a, p. 28). De esta manera, para determinar, como nos pedía Putnam en (1975c), si un término “T” tiene el mismo significado cuando es usado por dos sujetos distintos (y si estos sujetos expresan el mismo concepto mediante su uso), sólo hace falta tomar en consideración las historias causales de estos hablantes y averiguar si dichas historias conectan a los hablantes con entidades o conjuntos de entidades esencialmente distintos. En caso afirmativo, debemos determinar los significados y conceptos relevantes como significados y conceptos diferentes.¹¹⁸

¹¹⁸ Como arguiremos más adelante, este retrato de la teoría semántica en la que, supuestamente, se basa el externismo psicológico originado a partir de las ideas de Putnam no es correcto desde el momento en que

Bilgrami examina los experimentos mentales de Putnam y Burge para sacar a luz su parentesco con esta “teoría de la referencia directa” (Bilgrami 1992a, p. 22). En lo que respecta a las propuestas del primero, Bilgrami percibe en la afirmación de Putnam de que el agua es H₂O y sólo H₂O el sesgo esencialista que caracteriza a la “teoría de la referencia directa”. Esta afirmación, recordemos, le lleva a rechazar que cualquier otro líquido que tenga una composición química distinta sea agua, a pesar de que comparta con ella todas sus propiedades macroscópicas. En el caso de Burge, su compromiso con el esencialismo se manifiesta en una de las premisas de su experimento mental sobre el concepto de artritis expuesto en (1979), en particular, la afirmación de que si una comunidad lingüística distinta de la nuestra aplicara el nombre de “artritis” no sólo a los padecimientos artríticos, sino también a ciertos trastornos musculares, entonces la enfermedad así denominada no será artritis, debido a que, para Burge, la artritis es esencialmente una dolencia de las articulaciones. De este modo, como las prácticas lingüísticas socialmente establecidas para el término “artritis” determinan una aplicación de este término a una enfermedad distinta en la situación contrafáctica, Bert carecerá en dicha situación del concepto de artritis.¹¹⁹

Analizada la presencia del esencialismo, Bilgrami se aproxima a los ejemplos del externismo ortodoxo para explicitar su compromiso con la visión del significado de la “teoría de la referencia directa”. Para empezar, este autor sostiene que ese compromiso resulta mucho más evidente en los experimentos de Putnam. Así, para afirmar que el término “agua” usado por Óscar₁ difiere en significado al ser usado por Óscar₂ se ofreció una justificación que descansaba en la diferente naturaleza del líquido así denominado y, consiguientemente, en las distintas historias causales de ambos hablantes. En el caso de las tesis de Burge, su parentesco con la “teoría de la referencia directa” podría quedar oculto a simple vista. El motivo es que el externismo social sí tiene en cuenta las creencias y los modos de representación conceptual de los sujetos, en la medida en que alude a los usos que nuestra comunidad determina (conceptualmente) para nuestros términos lingüísticos. Ahora bien, lo interesante para Bilgrami es que Burge legitima la atribución de la noción de artritis a un hablante, Bert, que carece de un

tenemos en cuenta que para este autor el significado de un término es un vector de cuatro elementos, siendo tres de ellos (los marcadores sintácticos, los marcadores semánticos y los estereotipos) componentes que los hablantes deben conocer para ser competentes en el uso del término. Se comprueba con esto que, al contrario de lo afirmado por Bilgrami, desde el “externismo ortodoxo” las representaciones cognitivas de estos hablantes son un factor que debe tomarse en consideración para realizar atribuciones de contenidos.

¹¹⁹ Este autor denomina visión “científico-esencialista” al externismo de Burge, mientras que el externismo de Putnam tiene un carácter “causal-esencialista” (Bilgrami 1992b, p. 389).

dominio completo de dicha noción. Esto es, Bilgrami quiere destacar que, si bien es necesario contar con las concepciones asociadas a una noción particular para establecer que un hablante posee dicha noción, estas concepciones pueden ser patrimonio de un grupo de hablantes distinto del sujeto que recibe la atribución. La razón por la que Burge se ve inducido a realizar esta atribución conceptual a Bert reside, según Bilgrami, en una idea que comparte con la “teoría de la referencia directa”, a saber, que basta con que un agente permanezca en las relaciones causales (no-cognitivas) apropiadas para que sus palabras tengan el significado que tienen y exprese con ellas los conceptos que expresa. En este caso, la relación consiste en la pertenencia a una comunidad lingüística determinada. Esta tesis, crucial para el experimento mental de Burge, da lugar a las atribuciones inconsistentes que Bilgrami enuncia en la primera premisa de su trilema. El problema consiste en que, por un lado, Burge concibe la artritis (esencialmente) como una afección de las articulaciones y, por otro, permite atribuir a Bert el mismo concepto que posee su médico, lo que lleva a decir de él que cree que padece en el muslo una enfermedad que afecta sólo a las articulaciones, algo que resulta claramente inconsistente. Por ello, decir que Bert posee el concepto de artritis cuando dice “creo que tengo artritis en el muslo” atenta contra la presuposición de que Bert es un sujeto racional, en cuanto que implica afirmar que Bert cree algo de la artritis que es contradictorio con la esencia de dicha enfermedad. El externismo ortodoxo de Burge, concluye Bilgrami, entra en conflicto con el Principio de Caridad al autorizar atribuciones inconsistentes como ésta.

En (2003) este autor ofrece, además, una versión modificada del experimento mental de la Tierra Gemela¹²⁰ para mostrar que las atribuciones inconsistentes también se encuentran presentes en el externismo ortodoxo de Putnam. Bilgrami nos invita a imaginar a un sujeto de la Tierra, Burt, quien profiere la oración “el agua no es H₂O”. Esta oración es intuitivamente falsa porque, según sabemos, la composición química del agua es H₂O. Ahora bien, el externismo ortodoxo no respeta dicha intuición, puesto que, de acuerdo con Bilgrami, Putnam debería sostener que la creencia de Burt no sólo es falsa, sino contradictoria. A este respecto, afirma que debemos decir de él que cree de una sustancia que es H₂O que no es H₂O (Bilgrami 2003, p. 26). La “teoría de la referencia directa” en la que Putnam se apoya defiende una especificación del concepto de agua en función de su extensión, es decir, en función de su aplicación al conjunto de

¹²⁰ Seriamente modificada, en realidad, ya que el experimento original de Putnam se sitúa en 1750, época en la que “H₂O” sería una expresión carente de sentido para cualquiera.

los ejemplares cuya composición química es H₂O. Si esto es así, el externismo ortodoxo sustentado en esa teoría permite “enchufar” (“to plug in”) una descripción como “sustancia que es H₂O” en nuestros informes de contenidos psicológicos que involucran el concepto de agua. Pero de ser esto cierto, las atribuciones autorizadas por el externismo ortodoxo harían que Burt se comportara como un *idiota lógico* al mantener creencias manifiestamente contradictorias.

Establecida la primera premisa del trilema, Bilgrami enuncia los dos recursos que tiene ante sí un partidario del externismo ortodoxo para minimizar el impacto de las atribuciones no caritativas que se desprenden de su teoría. Por un lado, podría optar por restringir el acceso que un sujeto tiene a sus propios conceptos. De esta forma, Bert tendría el concepto de artritis al precio de desconocer, en parte, dicho concepto e ignorar, en consecuencia, que es esencialmente el concepto de una enfermedad de las articulaciones. No obstante, asumiendo la intuitiva plausibilidad del acceso privilegiado a los propios contenidos mentales, esta salida sólo redundaría en perjuicio del externismo. Por otro lado, de no estar dispuestos a renunciar ni al externismo ni al acceso privilegiado, sólo nos quedaría un último recurso, a saber, bifurcar el contenido. Si eligiéramos esta tercera salida, tendríamos que sostener que existen dos tipos de contenido, en primer lugar, un contenido en sentido restringido, internamente determinado y con respecto al cual poseemos pleno acceso epistémico, y, en segundo lugar, un contenido en sentido amplio, individuado a partir de factores externos que pueden ser desconocidos. Sin embargo, Bilgrami afirma que sería deseable una concepción unificada del contenido y en caso de que un externista ortodoxo tomara la opción bifurcacionista para compatibilizar su tesis con el acceso privilegiado estaría, en realidad, renegando del propio externismo y limitando seriamente el alcance de sus conclusiones.¹²¹

El trilema de Bilgrami queda establecido de la siguiente manera. O bien el partidario del externismo ortodoxo admite que su visión del contenido lleva a atribuciones inconsistentes, o bien limita el acceso epistémico a los propios pensamientos para evitar dichas inconsistencias, o bien bifurca el contenido, reservando la determinación externista para el contenido en sentido amplio y el acceso privilegiado para el contenido en sentido restringido.

¹²¹ Burge da la razón a Bilgrami en este sentido y rechaza la lectura bifurcacionista de los experimentos mentales de la Tierra Gemela (Burge 2007, pp. 11 y ss.). Sin embargo, a pesar de estas objeciones, existen desarrollos bifurcacionistas interesantes y dignos de ser tomados en consideración. Cf. Acero (1995).

Una vez expuesto el contenido del trilema, Bilgrami ofrece su propia comprensión de las tesis externistas, cuyo contenido viene recogido en la denominada RESTRICCIÓN C. Dicha restricción establece lo siguiente:

RESTRICCIÓN C: Al fijar un concepto externamente determinado de un agente, debemos hacerlo buscando las preferencias indéxicamente formuladas del agente que expresan contenidos indéxicos que contienen ese concepto y escogiendo después el determinante externo del concepto que está en consonancia con otros contenidos que han sido ya fijados para el agente.¹²²

Esta restricción impone una cierta metodología para realizar las atribuciones conceptuales. Por un lado, al atender a los componentes indéxicos de las preferencias de un hablante, podremos determinar qué conceptos está poniendo en ejercicio a través de su vinculación con los objetos e individuos (externos) sobre los que se aplican. La comprensión del externismo por parte de Bilgrami, recogida en esta primera parte de la RESTRICCIÓN C, constituye para este autor la condición de posibilidad de la publicidad del pensamiento y del acceso de la mente al mundo externo.¹²³

Sus propuestas, como vemos, tienen un claro parentesco con las de Davidson en dos aspectos. En primer lugar, ambos autores legitiman la identificación externista de los contenidos mediante argumentos de corte trascendental. Así, para estos autores la fijación externista de nuestros conceptos es lo que hace posible hechos bien probados, tales como el carácter público de las atribuciones mentales de los sujetos o el aprendizaje del lenguaje en tanto que lenguaje acerca de un mundo compartido.¹²⁴ En segundo lugar, ambos piensan que suscribir el externismo es la única opción para tener derecho a usar un cierto vocabulario en nuestras atribuciones de contenidos, a saber, un vocabulario acerca de objetos, individuos y propiedades externas. Bilgrami alega, en este sentido, que el uso de un vocabulario sobre elementos externos a los hablantes en la fijación de sus conceptos no sólo está respaldado por nuestras prácticas ordinarias en este sentido, sino que constituye el único modo posible de realizar atribuciones mentales. Ahora bien, el segundo aspecto de la RESTRICCIÓN C indica que la fijación conceptual a partir de factores externos debe dar lugar a atribuciones de conceptos que se encuentren en consonancia con otros conceptos y contenidos ya atribuidos a dicho hablante. Esta observación pretende ser el elemento diferenciador de la propuesta externista de

¹²² Cf. Bilgrami (1992b), p. 382 y (1992a), pp. 5 y ss.

¹²³ Bilgrami (1991a), pp. 200 y ss.

¹²⁴ Cf. Bilgrami (1992a), pp. 200 y ss.

Bilgrami con respecto a las tesis de Putnam y Burge, puesto que, de acuerdo con el trilema propuesto, el externismo ortodoxo defendido por estos dos autores lleva a la atribución de pensamientos inconsistentes.

Es necesario explicar otro aspecto de la metodología de las atribuciones conceptuales defendida por Bilgrami, un aspecto orientado, también, a respetar al máximo la presunción de racionalidad de los sujetos que reciben las atribuciones. La coherencia que se exige entre las atribuciones conceptuales y el resto del repertorio conceptual poseído por un hablante se ve reforzada por la distinción que efectúa Bilgrami entre dos niveles de la atribución de conceptos: el nivel agregativo y el nivel local.¹²⁵ El *nivel agregativo*, en el que operan las teorías del significado, es el ámbito en el que se sitúan todas aquellas creencias que un agente asocia con un concepto determinado. Ahora bien, en la medida en que es prácticamente imposible encontrar a dos hablantes que posean exactamente las mismas creencias asociadas con un concepto, también lo será encontrar a dos hablantes que compartan un concepto en el nivel agregativo. Por ello, no es en este nivel, sino en el *nivel local* aquél en el que tienen cabida nuestros informes sobre los pensamientos de terceros. Para este autor, las distintas localidades (situaciones o contextos) determinan o “destilan” qué creencias resultan relevantes para atribuir un determinado concepto a un sujeto. Por ejemplo, pensemos en una localidad como la siguiente: Pilar entra en una joyería porque quiere comprar un anillo de diamantes. Ella ignora que el diamante es un alótropo de carbono, que tiene muchas aplicaciones industriales debido a su gran dureza o que los diamantes más duros del mundo se encuentran en Nueva Gales del Sur. Por su parte, el joyero que atiende a Pilar tiene esas creencias y muchas otras acerca de los diamantes, lo cual significa que en el nivel agregativo existen grandes diferencias entre los conceptos de Pilar y los del comerciante. Aun así, las creencias de nuestra protagonista parecen ser suficientes para poder atribuirle el concepto de diamante en esta localidad, dado que no es necesario que un hablante posea creencias demasiado sofisticadas sobre los diamantes para decir de él que tiene el deseo de adquirir un anillo de diamantes.

Una vez que hemos explicado tanto el argumento de Bilgrami contra el externismo ortodoxo como las propuestas de este autor para elaborar un externismo respetuoso con el Principio de Caridad, conviene examinar el alcance de sus objeciones, así como la calidad de su propia contribución. En primer lugar, debemos indicar qué

¹²⁵ Cf. Bilgrami (1992b), pp. 386-387 y (1992a), p.141.

diferencias existen entre su comprensión de la “teoría de la referencia directa” y la caracterización de dicha teoría que hemos ofrecido en el primer capítulo de nuestro trabajo. Según dijimos, existen dos desarrollos distintos de la Nueva Teoría de la Referencia, uno que concierne al valor semántico de ciertas expresiones lingüísticas y otro que se ocupa de la determinación de la referencia de las mismas. Según la caracterización que Bilgrami ofrece de esta teoría, sus intereses tienen que ver, más bien, con este segundo aspecto. No obstante, la forma en que este autor describe la presencia de esta vertiente de las Nuevas Teorías de la Referencia en las propuestas de Putnam y Burge no parece ajustada al contenido de las tesis de estos dos autores. Con respecto a Putnam, este autor muestra en (1975c) que la determinación de la referencia de nuestros términos de género natural no se lleva a cabo por un mecanismo puramente conceptual. Esto es, la intensión de un término de esa clase, entendida como concepto o como entidad plenamente cognoscible por la mente individual, no es suficiente por sí misma para determinar una extensión correspondiente. Ahora bien, eso no significa que Putnam considere que para identificar el significado de un término de género natural “T” que figura en las preferencias de S la única consideración relevante sea si S ha entrado en contacto causal con la extensión de “T”, algo que se desprende de la caracterización que ofrece Bilgrami de la “teoría de la referencia directa”. De hecho, en Pessin/Goldberg (1996, p. xxi), encontramos un comentario de Putnam que desautoriza la lectura que Bilgrami ha realizado de sus propuestas en semántica:

“Desafortunadamente, más de un autor ha leído este ensayo [*‘El significado de ‘significado’*”] pasando por alto el hecho de que allí se afirma que hay factores adicionales involucrados en el significado. *‘El significado de ‘significado’* también contiene una incipiente teoría de la semántica de los términos de género natural y esa teoría *no* afirma que su significado esté enteramente determinado por la naturaleza de los objetos en su extensión (haya sido descubierta o no esta naturaleza por la investigación científica), sino que sostuve y sigo sosteniendo que está *parcialmente* determinado de ese modo”.

En efecto, recordemos que, para Putnam, la extensión sólo constituye un ingrediente del significado de un término de género natural “T” y que el resto de los componentes del vector significado de “T” (marcadores sintácticos, semánticos y estereotipos) son componentes que han de dominar aquellos sujetos competentes en el uso de “T”. Por tanto, cuál sea la extensión de “T” cuando es usado por S constituye, por supuesto, una condición necesaria para determinar el significado de “T”, pero la

extensión del término no lo es todo con respecto a su significado y, especialmente, con respecto a nuestro dictamen sobre la competencia de S en el uso de “T”.

En lo que concierne a Burge, es frecuente encontrar en sus escritos la idea de que el contexto socio-lingüístico de un hablante determina, *en parte*, qué conceptos expresa dicho hablante mediante el uso de sus términos. Como pudo apreciarse en la descripción del experimento mental de (1979), Bert no es un completo ignorante con respecto al concepto de artritis. Él *sabe* que es una enfermedad, que padece artritis desde hace años y que es preferible tener artritis a tener cáncer y, de acuerdo con ello, acude al experto indicado para resolver sus dudas. Según Burge, las prácticas lingüísticas de la comunidad de un agente son relevantes a la hora de atribuir un concepto u otro a dicho agente, *pero no son lo único relevante*. Desde el externismo social no es permisible atribuir el concepto (socialmente constituido) de artritis a Jesús, quien sólo sabe decir de la artritis que ayer desayunó artritis con patatas.¹²⁶ La ignorancia con respecto al uso del término “artritis” que se trasluce en esa afirmación es distinta de la ignorancia relativa de Bert con respecto al uso de “artritis” cuando profiere la oración “sufro artritis en el muslo”. O por lo menos es tan distinta como para considerar que una atribución del concepto de artritis violentaría la perspectiva cognitiva de Juan, mientras que decir que Bert tiene creencias en torno a la artritis sólo supondría una mejora caritativa del dominio que este hablante tiene del concepto de artritis.

Al afirmar que el externismo ortodoxo se apoya en una teoría de la referencia que prescinde de toda mediación conceptual para determinar cuál es el significado de los términos que usa un hablante y qué conceptos expresa mediante su uso, Bilgrami está hablando de un externismo que no es el que está basado en las ideas semánticas de Putnam ni el propuesto por Burge, puesto que, como vemos, ninguno de estos desarrollos prescinden del material cognitivo de los sujetos de las atribuciones semánticas y conceptuales.¹²⁷

¹²⁶ Descartamos un uso irónico o metafórico del término “artritis” por parte de este hablante.

¹²⁷ Por otro lado, cabe cuestionar que el esencialismo involucrado en las tesis de Putnam y Burge permita las atribuciones inconsistentes que Bilgrami denuncia. En este sentido, el hecho de que la artritis sea, esencialmente, una enfermedad de las articulaciones y que el agua sea, esencialmente, H₂O, no implica que todo hablante competente en el uso de los términos “artritis” y “agua” pueda acceder a la esencia de estos objetos en virtud de un conocimiento puramente conceptual. Dicho de otra manera, el dominio de estos conceptos ni requiere ni conlleva conocer la esencia de aquello a lo que se aplican. De hecho, los partidarios de la teoría de la referencia directa afirman que una oración de identidad como “el agua es H₂O” expresa una verdad necesaria *a posteriori* (Kripke, 1972, p. 128; Soames 2002, p. 15; Salmon 2003, p. 112). Esa identidad es necesaria porque el agua es esencialmente H₂O pero, debido a que la verdad de dicha identidad sólo puede conocerse *a posteriori*, ese establecimiento exige emprender indagaciones

Bilgrami, no obstante, intenta reforzar su argumentación planteándose qué clase de objeciones podrían ofrecer los partidarios del externismo ortodoxo para impedir el establecimiento de la primera premisa de su trilema. En primer lugar, considera la posibilidad de optar por una *teoría desentrecomilladora de los conceptos*, esto es, una teoría que defiende que para poseer el concepto de artritis basta con usar el término “artritis”, sin que se requiera que el hablante sea capaz de ofrecer explicaciones adicionales en cuanto a su uso de este término.¹²⁸ Esta teoría en principio impide establecer aquellas redescpciones inconsistentes de las creencias que figuran en la primera premisa del trilema (“Bert cree que tiene en su muslo una enfermedad que sólo afecta a las articulaciones”). Bilgrami replica que una estrategia de este tipo, que prescindiera de toda representación cognitiva asociada a los conceptos, sólo conseguiría convertirlos en entidades inefables. Ahora bien, en nuestra opinión los partidarios del externismo ortodoxo no tienen por qué optar por una teoría de este signo, ya que tanto Putnam como Burge sostienen que la determinación externista contribuye sólo en parte a la individuación de nuestros conceptos y que es necesario contar también con mediaciones cognitivas por parte del hablante que es sujeto de la atribución conceptual.

Bilgrami se enfrenta a un segundo recurso que el externismo ortodoxo podría ofrecer para cortocircuitar el establecimiento del trilema, a saber, *la distinción que plantea Burge entre conceptos y explicaciones conceptuales*. De acuerdo con Burge, las “explicaciones conceptuales” son las especificaciones reflexivas que podría proporcionarnos un agente acerca del valor cognitivo de un determinado concepto. En estas especificaciones un hablante podría hacer uso de otros conceptos relacionados y de inferencias a partir de experiencias perceptivas (Burge 1989a, p. 284). Según Bilgrami, tras esta distinción se oculta una bifurcación de contenido *ad hoc*, basada en una bifurcación subyacente de conceptos. De acuerdo con este autor, “se nos dice que hay dos nociones de conceptos 1) conceptos propiamente dichos, dados por la referencia externista y 2) conceptos en el sentido de explicaciones conceptuales que el agente puede articular bajo reflexión” (Bilgrami 1992b, p. 370). Sin embargo, no nos parece que Bilgrami esté haciendo justicia a la propuesta de Burge, ya que sólo podríamos aislar la primera categoría de conceptos si hubiéramos aceptado que es posible realizar atribuciones conceptuales atendiendo únicamente a las relaciones causales que un

empíricas. Si exigiéramos el conocimiento de las esencias de los géneros naturales para atribuir el dominio del término general que los designa, entonces estaríamos proponiendo un estándar de dominio extremadamente exigente.

¹²⁸ Bilgrami (1992a), pp. 43 y ss. y (1992b), pp. 373 y ss.

hablante ha establecido con la extensión del concepto, aunque careciera por completo de información asociada al mismo y fuera, por ello, incapaz de proporcionar explicación alguna sobre su uso del término que lo expresa. Sin embargo, las consideraciones previas nos impiden compartir esta descripción del externismo ortodoxo, puesto que los vínculos causales entre un sujeto y la extensión de sus conceptos (o el uso socialmente determinado de los mismos) no son condición suficiente para las atribuciones conceptuales. Burge exige, también, que todo hablante que amerite la atribución de un determinado concepto sea capaz de racionalizar la forma en que lo usa (Burge 1986c, p. 268) y, además, no parece existir en Burge una bifurcación propiamente dicha entre conceptos y explicaciones conceptuales. La pretensión de este autor consiste, simplemente, en explicar el hecho intuitivo de que podemos encontrar dos sujetos capaces de ofrecer diferentes articulaciones conceptuales si se les pidiera una explicación acerca de lo que entienden por un *mismo* concepto.

En tercer lugar, Bilgrami se enfrenta a una última estrategia que podría seguir el partidario del externismo ortodoxo para afrontar su trilema. Este recurso consiste en argüir que un hablante puede asociar con un cierto término *descripciones metalingüísticas* en las que encomienda a otros hablantes la tarea de determinar la aplicación del término. De ser así, no tendríamos por qué conocer todos los criterios de aplicación de los términos que usamos, puesto que podríamos servirnos del conocimiento que otros tienen al respecto. El dominio que un agente tiene sobre el concepto que expresa haciendo uso de uno de sus términos se vería “mejorado” gracias a que en esas descripciones se apela a otros hablantes más competentes en ese sentido (Bilgrami 1992b, p. 372). Estas descripciones tendrían la forma “artritis es lo que quiera que los expertos de mi comunidad denominen *artritis*”. Según Bilgrami, si un defensor del externismo ortodoxo tomara esta vía, esto resultaría fatal para su propia posición en tanto que externista. La razón es que nuestros conceptos recibirían entonces una individuación puramente interna, determinada exclusivamente por las descripciones metalingüísticas asociadas con los términos que expresan tales conceptos.¹²⁹ Sin

¹²⁹ No obstante, cabe realizar una observación a propósito de esta objeción de Bilgrami. Supongamos que un partidario del externismo aceptara que la determinación conceptual tiene lugar, en parte, por la presencia de descripciones metalingüísticas en las que se apela a los expertos de su entorno lingüístico. La pregunta que deberíamos hacernos es si el proponente de esta idea tiene derecho a seguir denominándose a sí mismo “externista”. Según lo vemos, este derecho existe desde el momento en que da cabida a elementos externos (los expertos relevantes), aunque los incorpore de manera mediata, utilizando para ello “elementos internos” como son las descripciones metalingüísticas. Ante la previsible objeción de que un autor internista como Searle podría ver reconocida su propia posición en la tesis descrita, responderemos que esto no supone despojar a esa tesis de su carácter externista (débil), sino que, más

embargo, pensamos que las conclusiones de Bilgrami no resultan del todo correctas. Para justificar esta afirmación vamos a examinar el experimento de Burge en torno a la artritis atendiendo a las observaciones de Putnam en (Pessin/Golberg 1996, p. xix). Así, si Bert y el Bert contrafáctico delegaran en los expertos de sus respectivas comunidades la tarea de determinar qué concepto están expresando mediante el uso del término “artritis”, los conceptos así determinados no serían idénticos. Y ello se debe al hecho de que uno y otro apelan a *distintos grupos de expertos*, es decir, a grupos de expertos situados en entornos distintos en un sentido relevante. La presencia de un componente indéxico en la descripción metalingüística “artritis es lo que quiera que los expertos de *mi* comunidad denominen *artritis*” prueba que esta clase de determinación es dependiente del entorno socio-lingüístico al que pertenece el hablante que asocia esa descripción con el término “artritis”. Desde el momento en que dos grupos de expertos utilizan distintos criterios para aplicar el término “artritis”, apelar a un grupo u otro no es indiferente, debido a que conlleva identificar de una u otra manera qué concepto expresamos al usar dicho término.

Las conclusiones que hemos podido establecer hasta el momento ponen de manifiesto que no es tan fácil sostener, como hace Bilgrami, que el externismo ortodoxo legitima atribuciones conceptuales inconsistentes. Sin embargo, queremos avanzar en nuestro examen preguntándonos, en primer lugar, si el externismo sujeto a la RESTRICCIÓN C que propone este autor no lleva aparejadas atribuciones conceptuales “mejoradas”, como las que se desprenden del externismo ortodoxo,¹³⁰ y, en segundo lugar, si la atribución conceptual en base a niveles descrita por Bilgrami responde a nuestras intuiciones sobre lo que significa dominar un concepto.

Para llevar a cabo la primera tarea, conviene confrontar el externismo *à la* Bilgrami con una versión del experimento mental que Putnam plantea en torno al uso de los términos “aluminio” y “molibdeno”.¹³¹ Recordemos que las premisas de dicho experimento retratan, en primer lugar, a un hablante de nuestro entorno, Óscar₁, que

bien, supone un reconocimiento por parte de los partidarios del internismo de que cualquier concepción razonable de la individuación de los contenidos mentales ha de dar cabida a elementos externos (mediatos) en estas tareas de individuación.

¹³⁰ Entendemos por atribución conceptual “mejorada” aquella que atribuye a un individuo la posesión de un concepto que sólo domina parcialmente, tal y como ocurre con Bert y la atribución del concepto de artritis en el experimento mental de Burge.

¹³¹ En efecto, el experimento que detallamos a continuación no es exactamente el experimento mental de Putnam que ofreció en (1975c), sino una versión del mismo más interesada en la determinación de los conceptos expresados por los términos “aluminio” y “molibdeno” que por la semántica de dichos términos.

tiene una serie de creencias verdaderas sobre el aluminio (cree que es un metal, que la mayoría de los enseres están fabricados de aluminio, que es más abundante que el molibdeno). Sin embargo, es incapaz de distinguir el aluminio respecto del molibdeno. En segundo lugar, pensemos en un entorno contrafáctico donde se llama “aluminio” al molibdeno y “molibdeno” al aluminio y donde, además, el molibdeno es tan abundante como (y tiene los mismos usos que) el aluminio en nuestro entorno. Óscar₂, un habitante de ese entorno posible, enteramente idéntico a Óscar₁ en sus aspectos intrínsecos, estaría dispuesto a asentir a preferencias como “el aluminio es un metal”, “el aluminio es más abundante que el molibdeno” y “la mayoría de nuestros enseres están fabricados con aluminio”, aun cuando, al igual que Óscar₁, no supiera distinguir entre estos dos metales. Lo que debemos preguntarnos es si el concepto que expresan Óscar₁ y Óscar₂ mediante el término “aluminio” al proferir “esta olla está hecha de aluminio” debe ser el mismo concepto atendiendo a la RESTRICCIÓN C. Respetando el lema de dicha Restricción, deberíamos correlacionar las preferencias indécicas de estos sujetos con los elementos de su entorno externo a los que se refieren. Ahora bien, mientras que la olla a la que se refiere Óscar₁ es de aluminio, la olla de la que habla Óscar₂ es de molibdeno y, por tanto, parece difícil decir que uno y otro están expresando el mismo concepto mediante el uso del término de género natural “aluminio”, a pesar de que ambos sean incapaces de establecer en qué se diferencian el aluminio y el molibdeno. Por tanto, el distinto carácter de los determinantes externos de las preferencias de estos agentes nos insta a fijar sus conceptos relevantes como conceptos distintos, tal y como defienden Putnam y Burge. Lo que resulta más interesante es que la propia RESTRICCIÓN C autoriza este resultado, es decir, un resultado que “mejora” el conocimiento que los hablantes tienen de sus propios conceptos, puesto que nos invita a distinguir entre las creencias que mantienen dos hablantes que no son capaces de establecer desde un punto de vista subjetivo dicha diferencia.¹³²

¹³² En nuestra opinión, resulta significativo que Bilgrami suela centrar sus críticas en los experimentos mentales sobre el concepto de artritis y el concepto de agua, donde los supuestos esencialistas son más evidentes. Y es significativo porque el esencialismo es una tesis muy controvertida y, como veremos en el último capítulo, existen buenos argumentos para ver en el XYZ un tipo de agua. Ahora bien, estamos de acuerdo con Crane en que lo único que hace falta para construir un experimento mental a favor del externismo es que los individuos que protagonizan estos experimentos se refieran a algo *distinto* aunque todo les parezca igual (Crane 2001, p. 123). Si uno piensa que carece de sentido hablar de esencias, no tiene por qué aceptar un experimento mental que gire en torno a determinantes conceptuales externos *esencialmente* distintos. Sin embargo, cabe desplazar las esencias como criterio de identidad hacia el acuerdo colectivo de los hablantes sobre dichos criterios. Por ello, pensamos que Bilgrami podría aceptar perfectamente un experimento en el que figuren dos determinantes acerca de los cuales nadie esté dispuesto a afirmar que se tratan de la misma entidad (como ocurre con el aluminio y el molibdeno).

Antes de dar por concluida esta sección, queremos preguntarnos qué consecuencias tiene la atribución en base a niveles propuesta por Bilgrami a la hora de proponer unos estándares intuitivos para otorgar a un sujeto competencia en el uso de un concepto. Con el fin de establecer un fondo de contraste, empezaremos exponiendo las tesis que Burge mantiene a este respecto. Según este autor, para atribuir un concepto a un sujeto bastaría con que reconociésemos en este sujeto la capacidad de llevar a cabo *racionalizaciones responsables* en torno a su uso de este concepto si alguien se lo solicitara. Supongamos que María dice “tengo gripe” y, a continuación, se le pide a este hablante explicar qué entiende por “gripe”. Si María fuera absolutamente incapaz de proporcionarnos la más mínima indicación sobre este asunto, si no supiera vincular ese concepto con otros como enfermedad, virus, congestión o fiebre, nadie consideraría que está expresando un pensamiento sobre esa enfermedad cuando dice “tengo gripe”. A falta de una racionalización responsable sobre su uso del término en cuestión, consideraríamos que está hablando caprichosamente e, incluso, si esa oración fuera verdadera, creeríamos que María se está limitando a repetir unas palabras que ha escuchado sin comprenderlas en absoluto. Una vez que hemos examinado el estándar de competencia mínima que Burge señala para la atribución de un concepto, podemos desplazarnos al extremo contrario, el de la máxima competencia. Según este autor, con respecto a un concepto dado existe un grupo de hablantes que dominan especialmente bien las condiciones de su aplicación; se trata de los expertos, un conjunto de hablantes que establecen los estándares de uso de los conceptos. Dichos estándares nos servirán para medir la competencia de los demás hablantes en el dominio de esos conceptos. Por último, entre la exigencia de racionalización mínima y el dominio conceptual experto, encontramos multitud de casos intermedios. De hecho, el propio Bert, quien protagoniza el experimento mental de Burge, se sitúa en este umbral de mediana competencia, debido a que, por un lado, tiene una serie de creencias verdaderas en torno a la artritis, pero, por otro, lleva a cabo aplicaciones incorrectas puntuales del término “artritis”. Lo que debemos retener es que para Burge el dominio de un concepto admite grados. Así, uno es más o menos competente en el uso de un término y, por tanto, uno tiene un conocimiento más o menos completo del concepto que expresa mediante su uso. Sin embargo, preguntar si un sujeto posee un concepto es plantear una cuestión que sólo admite dos respuestas, sí o no, aunque quepa matizar hasta qué punto lo domina en caso

de poseerlo. Por descontado, siempre podemos rectificar las atribuciones conceptuales realizadas si las evidencias posteriores al momento de la atribución nos invitan a ello.¹³³

Dicho esto, lo que nos interesa señalar es la diferencia que existe entre las pautas de atribución conceptual que Burge propone y la metodología en base a localidades que defiende Bilgrami a este respecto. Recordemos que dichas localidades no son más que los contextos en los que dictaminamos si un agente posee un concepto. Estas localidades “destilan” cuáles son las creencias que un hablante debe poseer para ser merecedor de la atribución. Por ejemplo, resulta razonable pensar que si Juan se lamenta por la pérdida de su alianza de plata no es necesario que tenga la creencia de que el nombre de ese metal en latín es “*argentum*”. Basta con que sepa, por ejemplo, que es un material relativamente valioso con el que se fabrican alianzas como la suya. Ahora bien, la atribución basada en localidades supone que, a diferencia de Burge, Bilgrami no establece un nivel mínimo de competencia en el dominio de un concepto a partir del cual pudiéramos decir que el hablante S posee ya el concepto C, sino que son las distintas localidades las que “destilan” sus propios niveles de competencia mínima. De esta manera, en una localidad L podría decirse que S posee el concepto C si tuviera las creencias que la localidad hace relevantes. Si S cambiara de localidad, y cambiaran, con ella, las creencias destiladas, el hablante S podría carecer en esta nueva localidad del concepto cuya posesión le atribuimos anteriormente sin problemas, en caso de que careciera de las creencias relevantes que “destila” la nueva localidad. Esto significa que para Bilgrami nunca podemos establecer de manera definitiva la posesión de un concepto por parte de un sujeto susceptible de situarse en localidades que “destilan” creencias mínimas diferentes.¹³⁴

Según pensamos, esta consecuencia de las tesis de Bilgrami podría tener resultados un tanto contraintuitivos y, para mostrarlo, examinaremos una versión modificada del experimento mental de Burge a la luz de la atribución conceptual en función de localidades. Pensemos en una primera situación (A) en la que Bert, el

¹³³ Imaginemos que María dice “tengo gripe” con unos signos de congestión más que evidentes. Es probable que ante esto nuestra primera reacción sea atribuirle un pensamiento acerca de la gripe. Pero si al pedir a María que nos explique lo que entiende por “gripe” ella fuera incapaz de ofrecernos la más mínima indicación al respecto, tendríamos derecho a rectificar nuestro juicio y negar que María posea, en realidad, el concepto de gripe.

¹³⁴ Esta idea es distinta de la autorización para rectificar nuestras atribuciones. Rectificar significa decir “creía que S poseía el concepto C, pero *en realidad* no lo posee”. La idea de Bilgrami es que no cabe hablar de un estándar válido para todos los contextos a partir del cual podamos decir que un individuo “en realidad” posee un concepto. Cambiar de localidad supone cambiar de estándar y por ello, es posible decir que en L S no posee C, mientras que en L’ sí lo poseía.

hablante del mundo actual, le comenta a su vecina (quien también padece artritis) la evolución de su enfermedad. En esta localidad, las creencias cuya posesión es necesaria para atribuir tanto a Bert como a su vecina el concepto de artritis son creencias muy básicas y poco numerosas y, teniendo en cuenta lo que Bert sabe sobre la artritis, parece que estamos autorizados a atribuir ese concepto a ambos sujetos. Ahora pensemos en una segunda situación (B). En ella, Bert acude a su médico, tal y como relata Burge, para exponer su temor de que la artritis se esté extendiendo a su muslo. Sin embargo, en nuestra versión el médico no llega a corregir el uso de Bert del término “artritis” y se limita a recetarle calmantes. Según Bilgrami, debemos negar que Bert posea el mismo concepto que su médico, ya que Bert carece de la creencia de que la artritis es una enfermedad que no afecta a los músculos, relevante en la localidad (B). La última localidad (C) de nuestro experimento es muy similar a la primera localidad (A). Al regresar de la consulta, Bert vuelve a encontrarse con su vecina y se desean mutuamente una mejoría en sus padecimientos artríticos. Al igual que en (A), las creencias sobre la artritis que (C) “destila” son muy elementales y de acuerdo con ello estamos autorizados a atribuir dicho concepto a Bert, puesto que, por hipótesis, Bert posee tales creencias.

Lo interesante de este juego de localidades es que nos permite mostrar cómo, desde el análisis de las atribuciones conceptuales propuesto por Bilgrami, Bert puede poseer un concepto en un momento y lugar determinados y “perder” dicho concepto al desplazarse a una localidad distinta. Decimos que Bert “pierde” su concepto de artritis, lo cual es muy distinto de rectificar la atribución realizada en (A) alegando que Bert nos ha dado muestras en (B) de un dominio insuficiente del concepto. No existe rectificación en la medida en que, para Bilgrami, Bert poseía de manera efectiva el concepto de artritis en (A), aunque en (B) ya no lo posea y sin que lo haya olvidado. Por tanto, si la localidad impone qué creencias son relevantes, un hablante puede poseer un concepto en una localidad, no poseerlo en una localidad posterior, volver a recuperarlo en una tercera localidad, y así sucesivamente, atendiendo a la diferencia en las creencias relevantes señaladas por cada una de estas localidades. Sin embargo, nos preguntamos hasta qué punto resulta razonable una metodología de atribuciones conceptuales que da lugar a estas pérdidas y recuperaciones de conceptos. En nuestra opinión, resulta más verosímil reconocer en un hablante (para todas las localidades) distintos grados de dominio de un mismo concepto, que afirmar, con Bilgrami, que un hablante puede poseer un concepto en una localidad, perderlo en una segunda localidad y recuperarlo en

una tercera, todo ello sin que haya mediado accidente psicológico alguno (*lapsus* o pérdidas de memoria).

No obstante, ha de decirse que las pautas de atribución conceptual basada en localidades ideadas por Bilgrami encajarían perfectamente en una concepción instrumentalista de lo mental. Es decir, si los conceptos y contenidos mentales que atribuimos a los hablantes fueran meros recursos explicativos destinados a volver inteligible su conducta, esas pérdidas y recuperaciones que se desprenden del modo de atribución que defiende este autor no tendrían nada de anómalo. Simplemente, en una determinada localidad el comportamiento de Bert cobraría mayor sentido si le atribuyéramos creencias sobre la artritis, mientras que en otras localidades su conducta podría explicarse mejor si le despojáramos de dicho concepto. En cualquier caso, el instrumentalismo mantendría que no hay nada que haga verdadera una atribución conceptual, y, por tanto, los informes sobre conceptos y contenidos constituyen un recurso que cada uno de nosotros usamos sin esperar una correspondencia de dichos informes con algún hecho. Ahora bien, la excelente sintonía entre el instrumentalismo y la atribución en función de localidades debería ser un problema para Bilgrami, puesto que este autor ha rechazado explícitamente una concepción instrumentalista de lo mental (1992a, pp. 235 ss.). Por todo ello, podríamos confrontar a Bilgrami con un dilema. O bien suscribe la atribución de conceptos y contenidos basada en localidades, con la consabida consecuencia de hablar de “pérdidas” y “recuperaciones” de conceptos, o bien se acoge a una visión instrumentalista de lo mental, en cuyo seno las “pérdidas” y “recuperaciones” conceptuales resultan inocuas.

5. Conclusiones

En este capítulo hemos examinado dos argumentos que comparten una misma finalidad, a saber, revelar la incompatibilidad entre el externismo psicológico y el acceso privilegiado a los propios contenidos mentales. El primero de ellos, formulado en la primera sección de este capítulo, se alimenta del hecho de que la localización externa al sujeto de las condiciones de identidad de sus contenidos mentales supone que dicho sujeto podría desconocer dichas condiciones. Si saber lo que pensamos requiere conocer la identidad del contenido y si dicho contenido depende de factores que no conocemos con inmediatez o autoridad, entonces la autoridad, inmediatez y notoriedad que presumiblemente caracterizan al auto-conocimiento se verían en peligro. El modelo

basado en la inclusión, propuesto por Burge y respaldado por importantes representantes del externismo, pretende hacer frente a esta objeción mostrando que el procedimiento por el que logramos obtener conocimiento de nuestros contenidos mentales requiere de elementos que nada tienen que ver con el conocimiento de las condiciones de individuación de nuestros contenidos, sean estas condiciones internas o externas a nosotros mismos. La propia conformación de los juicios reflexivos básicos o juicios de tipo *Cogito* produce por sí misma la garantía de su efectuada exitosa, debido a que los juicios reflexivos heredan, reutilizan o incluyen lo juzgado como su propio contenido. Por ello, dice Burge, aunque Óscar₁ sufriera un traslado inadvertido a la Tierra Gemela y permaneciera en ese entorno el tiempo suficiente para que cambiara el concepto que expresa mediante el uso del término “agua”, este sujeto siempre sabría lo que cree en el momento de pensar en sus creencias.

Sin embargo, comprobamos que Burge admite que carecemos de la capacidad de reconocer en nosotros mismos cambios conceptuales como los que, según las premisas externistas, se producen en los casos de traslado. Nuestra intención es abordar en detalle el planteamiento incompatibilista derivado de los casos de traslado en el próximo capítulo, pero podemos adelantar que el núcleo del debate tendrá que ver con las consecuencias epistemológicas que se derivan de nuestra incapacidad para reconocer cambios conceptuales de esa clase. Por otro lado, hemos podido percibir que el cuestionamiento de la comprensión perceptiva de la introspección es una constante en los filósofos que respaldan el modelo basado en la inclusión. Mientras que estos autores ven en el externismo una excelente oportunidad para criticar los supuestos cartesianos que sustentan esa visión de la introspección, los partidarios del incompatibilismo sólo reconocerán en esta crítica la necesidad de diseñar planteamientos más elaborados, en los que no figuren metáforas sobre la “captación” de pensamientos, considerados como objetos presentes al ojo de la mente.

Por lo que respecta al segundo argumento analizado, el trilema de Bilgrami, hemos intentado mostrar que se basa en una comprensión excesivamente simplificada de los supuestos semánticos del denominado por él “externismo ortodoxo” y que, además, el externismo alternativo que propone este autor se encuentra comprometido con una visión instrumentalista de los conceptos y contenidos mentales. Sin embargo, es preciso reconocer que se trata de un planteamiento interesante y elaborado, que saca a relucir cierto aspecto polémico de las propuestas externistas de Putnam y Burge, a saber, la permisividad a la hora de atribuir a los sujetos conceptos sobre los que tienen un

dominio defectuoso. Ya el propio Davidson denunciaba que las tesis de Burge sobre este punto tienen un parecido abrasador con la falta de acceso privilegiado a nuestros propios pensamientos. Así, se dice, si podemos atribuir a un hablante un concepto que sólo domina en parte, estaremos diciendo que este hablante desconoce algunas de las condiciones de aplicación de los conceptos que posee. De nuevo, mientras que Davidson y Bilgrami consideran que esta objeción es suficiente para mostrar que el denominado por él como “externismo ortodoxo” atenta contra el privilegio epistémico de primera persona, Burge ve en esta réplica el fantasma de la comprensión perceptiva de la introspección y, denuncia que los conceptos no son “cosas” que uno posea, sino marcas de habilidades más o menos perfeccionadas (Burge 2003a, p. 277). A nuestro modo de ver, los partidarios del externismo psicológico basado en las ideas semánticas de Putnam y el anti-individualismo de Burge deberían reforzar este aspecto de sus propuestas, demostrando que las atribuciones que “mejoran” el dominio que los sujetos tienen de sus propios conceptos constituyen una práctica usual y perfectamente respetuosa con el Principio de Caridad, más bien que una violación de dicho Principio.

Capítulo 3: Los casos de traslado

En el presente capítulo nos ocuparemos de los argumentos basados en los casos de traslado, una estrategia incompatibilista diseñada para demostrar *de una manera indirecta* la imposibilidad de conjugar el externismo y el acceso privilegiado a los propios contenidos. Destacamos el carácter indirecto de esta argumentación para establecer un contraste con respecto al planteamiento incompatibilista simple. Según vimos, este planteamiento ponía en juego una confrontación inmediata entre los rasgos del conocimiento en primera persona y las tesis externistas; el resultado perseguido con dicha confrontación era demostrar la dificultad de preservar la autoridad, inmediatez y notoriedad propias del acceso epistémico a nuestros propios contenidos mentales cuando la identidad de estos contenidos depende, como afirma el externismo, de elementos que podemos ignorar. Los casos de traslado, por su parte, toman en consideración una extensión del escenario de la Tierra Gemela para concluir que la individuación externista da lugar a consecuencias implausibles tanto en el ámbito del conocimiento de los contenidos rememorados como en las inferencias realizadas a partir de contenidos adquiridos en diferentes entornos. Estos casos, que adoptan la forma de una reducción al absurdo del externismo, pretenden poner de manifiesto que nuestras intuiciones sobre el comportamiento de la memoria o sobre el carácter *a priori* de aquellas inferencias que realizamos tomando nuestros propios contenidos como premisas son intuiciones incompatibles con las tesis externistas y, puesto que no estamos dispuestos a renunciar a tales intuiciones, entonces sólo nos queda, según el incompatibilista, rechazar el externismo.

Existen buenas razones para que un partidario del incompatibilismo adopte una estrategia indirecta como la de los casos de traslado. Por una parte, las formulaciones del planteamiento simple suelen apoyarse, como vimos, en una consideración perceptiva del conocimiento introspectivo que reifica en exceso los contenidos que “ve” nuestra mente. Sin embargo, ya comprobamos que esta concepción del acceso privilegiado constituye una imagen de la mente tan simplista como cuestionable. Por otra parte, el modelo basado en la inclusión, propuesto por Burge y adoptado por externistas de gran relevancia, asegura el carácter auto-verificativo del conocimiento que tenemos de nuestros pensamientos *presentes*. De acuerdo con este modelo, la conformación de juicios reflexivos tiene lugar mediante un proceso que garantiza que los pensamientos

juzgados sean, a su vez, los contenidos de tales juicios. La conclusión derivada de este modelo es que el conocimiento de nuestros propios pensamientos en los casos más básicos es infalible, independientemente de la teoría de la individuación del contenido que uno suscriba.

Dicho esto, se entenderá que las iniciativas incompatibilistas más prometedoras hayan optado por abrir nuevas vías. La sospecha común a todas ellas es que la clase de acceso privilegiado que preserva el modelo basado en la inclusión no es el acceso privilegiado *tout court*. El propósito de los incompatibilistas al plantear los casos de traslado consiste en mostrar que a pesar de que la inclusión reflexiva nos permite blindar el acceso privilegiado a los pensamientos presentes, no es una estrategia que nos capacite para acceder, sin pérdida, a todos nuestros contenidos. Los casos de traslado son vistos como el mejor modo de explotar las consecuencias de esta laguna en las capacidades cognitivas de primera persona y cuestionar, con ello, la viabilidad del propio externismo.

El itinerario de este capítulo es el siguiente. En primer lugar, tomaremos en consideración el caso de traslado que Boghossian desarrolla en (1989). Este argumento parte de la suposición de que el externismo conlleva que un hablante que cambia inadvertidamente de entorno sufrirá un cambio en su repertorio conceptual igualmente inadvertido. Como resultado de ese cambio, dicho hablante no podrá acceder a los contenidos mentales que adquirió en su entorno pasado. A partir de esto, y supuesto que dicha incapacidad no se debe a un fallo en la memoria del sujeto, Boghossian argumenta que la única explicación posible es que, dado el externismo, este sujeto tampoco conocía los contenidos mentales rememorados en el momento de adquirirlos. Una vez expuesto el primer caso de traslado, discutiremos las diferentes respuestas que los partidarios del externismo han ofrecido para restar plausibilidad a este argumento. En la segunda sección de este capítulo emprenderemos el examen de una nueva versión de los casos de traslado que Boghossian presentó en sus trabajos de (1992) y (1994). El punto de partida de este argumento establece que, dado el externismo, la víctima de traslado es incapaz de discriminar entre sus contenidos mentales presentes y pasados y, llevado por esa falta de habilidades discriminativas, podría cometer errores en su razonamiento impropios de un sujeto caritativamente racional. La conclusión que Boghossian extrae de esta situación es la incompatibilidad entre el externismo y el Principio de Caridad, un principio básico que rige nuestras prácticas ordinarias de atribuciones de contenidos. Después de examinar esta segunda versión de los casos de traslado, abordaremos las

diferentes estrategias que se han ideado desde el compatibilismo para bloquear este nuevo argumento. Dedicaremos la tercera sección a discutir qué clase de relevancia es conveniente asignar a los casos de traslado y, finalmente, explicaremos qué concepción del externismo es la más adecuada, en nuestra opinión, para esquivar con éxito el desafío incompatibilista que los casos de traslado ponen en escena.

1. *Casos de traslado y memoria*

Como se ha indicado, esta sección está dedicada a la discusión del argumento incompatibilista que Boghossian presentó en (1989). Este trabajo pretende establecer que nuestra falta de habilidad para discriminar entre algunos contenidos pasados y presentes, carencia que Burge admite como una consecuencia del externismo en los casos de traslado, podría volverse en contra del modelo basado en la inclusión y, en esta medida, llegar a socavar el conocimiento de los contenidos mentales presentes, que es, según vimos, el tipo de conocimiento en primera persona cuya infalibilidad quedaba garantizada por la inclusión reflexiva.

Boghossian plantea su argumento a partir de una descripción de un caso de traslado muy parecida a la que propuso Burge en (1988a). Boghossian nos pide que imaginemos que Óscar₁, el habitante de la Tierra, es trasladado mientras duerme o está inconsciente a la Tierra Gemela. Es lógico suponer que este hablante no advertirá que dicho traslado ha tenido lugar, dado que, por hipótesis, la única diferencia que existe entre ambos entornos escapa a su conocimiento. Nuestro protagonista ignora, en consecuencia, que el líquido que corre por los ríos y lagos de la Tierra Gemela y que tanto él como los habitantes de su nuevo entorno denominan “agua” no es agua, de acuerdo con la concepción de los géneros naturales que sustenta este experimento mental. Boghossian sostiene que, según las pautas externistas, cuando Óscar₁ profiera la oración “el agua es símbolo de purificación” deberíamos atribuirle *a partir de ahora* una creencia distinta de aquella que le habríamos atribuido si dicha preferencia hubiera tenido lugar tiempo atrás en la Tierra.

Ahora bien, si atendemos al modelo basado en la inclusión, el proceso de formación de los pensamientos reflexivos *en presente* garantiza que cualquiera que sea el pensamiento de primer orden, éste constituirá, a su vez, el contenido del pensamiento de segundo orden. De este modo, si el traslado se traduce en una variación de las condiciones de individuación del contenido del pensamiento expresado mediante la

preferencia “creo que el agua es húmeda” – variación que conlleva que el pensamiento verse acerca del agua₂ y no sobre el agua –, un traslado inadvertido supondrá, consiguientemente, que el pensamiento reflexivo expresado mediante la preferencia “pienso que creo que el agua es húmeda” sea, asimismo, un pensamiento acerca del agua₂. Los compatibilistas pueden servirse de este modelo para garantizar que Óscar₁ sabía en t₁ (tiempo en el que residía en la Tierra) qué es lo que estaba pensando en t₁ y, además, que este hablante sabe en t₂ (tiempo en el que se ha producido el traslado) qué es lo que está pensando en ese instante t₂. No obstante, a pesar de que Burge considera que la inclusión reflexiva es suficiente para preservar el acceso privilegiado de la víctima de traslado, concede que la confrontación entre el externismo y la situación de traslado pone en evidencia un punto ciego en el conocimiento que Óscar₁ tiene de su vida mental.¹³⁵ En particular, este sujeto no sabe en t₂ que existe una diferencia entre los pensamientos que mantenía en t₁ y los que mantiene actualmente. Este es, precisamente, el punto de apoyo de las reclamaciones incompatibilistas de Boghossian, quien pretende demostrar que la ignorancia de Óscar₁ en relación al cambio que se ha producido en su propio repertorio conceptual socava el acceso privilegiado a sus propios contenidos mentales.

El hecho de que la inclusión reflexiva sólo garantice la infalibilidad en el autoconocimiento si los pensamientos de primer y segundo orden son “absolutamente coincidentes” en el tiempo (Boghossian 1989, p. 21) constituye una limitación de la que Boghossian se sirve para formular el siguiente argumento:

“Lo siguiente podría parecer una perogrullada sobre la memoria y el conocimiento: si S sabe que p en t₁, y si (algún tiempo después) en t₂, S recuerda todo lo que sabía en t₁, entonces S sabe que p en t₂. Ahora, preguntemos: ¿por qué S no sabe hoy si su pensamiento de ayer era un pensamiento sobre el agua o sobre el agua₂? La perogrullada insiste en que hay sólo dos posibles explicaciones: o bien S lo ha olvidado o bien nunca lo supo. Pero seguramente el hecho no es un fallo de la memoria [...] La única explicación, me atrevo a sugerir, para que S no sepa mañana lo que dice saber hoy, no es que lo haya olvidado, sino que nunca lo supo. Los juicios auto-verificativos de Burge no constituyen conocimiento genuino” (Boghossian 1989, p. 23).

¹³⁵ Cf. Burge (1988a), pp. 652 y ss. y (1988b), p. 213.

Vamos a examinar este razonamiento utilizando una estructuración del mismo que se ha convertido en canónica:¹³⁶

(M1) Si S no olvida nada entre t_1 y t_2 , lo que sabe en t_1 debe saberlo en t_2 .

(M2) S no olvida nada entre t_1 y t_2 .

(M3) S no sabe que P en t_2 .

Por tanto, S no sabe que P en t_1 .

El establecimiento de la primera premisa no debería ser problemático, según nos dice Boghossian, puesto que sólo constituye una “perogrullada” acerca del funcionamiento de la memoria. De este modo, todo el mundo esperaríamos que un hablante continúe teniendo acceso epistémico en el presente a un contenido mental adquirido en el pasado a menos que lo haya olvidado. La segunda premisa afirma que Óscar₁ no ha olvidado en la Tierra Gemela los contenidos mentales que tenía en la Tierra. La tercera premisa, por su parte, establece que en la Tierra Gemela (t_2) Óscar₁ no sabe qué piensa en ese entorno (t_2). Es decir, que al darse un cambio en los factores externos que, de acuerdo con el externismo, contribuyen a individuar los conceptos y contenidos de nuestro protagonista, Óscar₁ no podrá mantener en su nuevo entorno los contenidos que mantenía en su planeta de origen. En conclusión, si nuestro protagonista no ha olvidado nada entre t_1 y t_2 , y si esto implica, además, que lo que sabía en t_1 también debería saberlo en t_2 y aceptamos como una consecuencia del externismo que en t_2 Óscar₁ no sabe qué pensaba en t_1 , entonces no queda más remedio que concluir que, en realidad, este sujeto no supo nunca qué es lo que pensaba en t_1 .

Antes de comenzar con nuestro examen de las réplicas compatibilistas, es conveniente destacar que la argumentación de Boghossian no se atiene estrictamente a la formulación inicial de los casos de traslado ideada por Burge. En primer lugar, recordemos, este último autor sostuvo que el cambio en los contenidos de Óscar₁ no se produciría inmediatamente después del traslado. Sin embargo, Boghossian habla de t_1 y t_2 en términos de “hoy” y “mañana”, lo que contraviene las indicaciones de Burge. Aunque podemos admitir que hablar en estos términos sólo constituye un recurso teórico destinado a exagerar las consecuencias escépticas en el auto-conocimiento derivadas del externismo, existe un segundo aspecto del relato original que Boghossian pasa por alto y que resulta un tanto más problemático. Se trata del supuesto que sostiene la premisa (M3) que establece que Óscar₁ no sabe que P en t_2 . Boghossian parece dar

¹³⁶ Cf. Bernecker (2004), p. 608; Brueckner (1997), p. 323; Burge (1998), p. 356; Goldberg (2003a), p. 243-244 y (1997), p. 214; Kraay (2002), p. 297; Ludlow (1995b), p. 308 y (2004), p. 419.

por hecho que el viaje al nuevo entorno conlleva una *pérdida* de los conceptos terrestres de Óscar₁. En efecto, para sostener que Óscar₁ no puede saber en la Tierra Gemela qué es lo que pensaba en la Tierra habría que justificar que en su nuevo entorno nuestro protagonista es incapaz de usar y de traer a la memoria el concepto terrestre. Sin embargo, en el relato original de Burge en (1988a) no se afirma tal cosa en ningún momento.¹³⁷

A lo largo de esta sección vamos a examinar las diferentes réplicas que los partidarios del externismo compatibilista han diseñado para mostrar que el argumento de Boghossian no es concluyente. Debemos señalar, eso sí, que la mayor parte de estas respuestas se ha centrado en la discusión de los supuestos sobre el funcionamiento de la memoria que mantiene Boghossian, algo que para algunos autores ha desviado el debate de su centro de gravedad.¹³⁸

Para abordar el examen de las respuestas compatibilistas, hemos optado por clasificar dichas respuestas atendiendo a qué premisa del argumento sobre la memoria es la que cuestionan. Así, comenzaremos con los argumentos de Brueckner en (1997) y de Ludlow en (1995b) y (2004) contra la premisa (M1). En segundo lugar, veremos qué estrategia ha elegido Tye en (1998) para cuestionar la premisa (M2). En tercer lugar,

¹³⁷ Como veremos, varios autores compatibilistas (entre los que encontramos al propio Burge) niegan que la pérdida conceptual sea la norma en los traslados.

¹³⁸ Cf. Goldberg (1997). En ese trabajo Goldberg ofrece una reformulación del argumento de Boghossian de (1989) que no apela al funcionamiento de la memoria de Óscar₁. Al hacerlo Goldberg pretende extraer la intuición que subyace a este argumento evitando discusiones sobre el funcionamiento de la memoria, sus funciones y clases. No obstante, en su reformulación introduce una variación sobre el relato del caso de traslado de Boghossian: Óscar₁ ha sido informado de que el traslado ha tenido lugar, aunque sin indicarle cuándo. Siendo W la oración “el agua es húmeda”, y S el protagonista del caso de traslado, Goldberg enuncia así su argumento:

- (1) Si S se auto-adscribe un pensamiento con una forma verbal W que es tal que,
 - a. Por sus propios medios, S sabe que hay más de una interpretación de W y
 - b. S no tiene disponible en el presente ninguna manera de seleccionar una de estas interpretaciones.

Entonces la auto-adscripción de S no cuenta como auto-conocimiento porque no es una identificación consciente del pensamiento en cuestión.

- (2) La auto adscripción del pensamiento de que P de S satisface (a) y (b).
- (3) Por tanto, la auto-adscripción de S no cuenta como auto-conocimiento porque no es una identificación del pensamiento en cuestión.

Según Goldberg, este argumento permite mostrar la incompatibilidad existente entre el externismo psicológico y un cierto tipo de auto-conocimiento que él mismo denomina de distintas maneras: “identificación deliberada de pensamientos auto-adscritos” (1997), “conocimiento discriminativo del contenido” (1999) y “auto-conocimiento conceptual” (2003a). La carencia de este conocimiento conlleva que la víctima de traslado es incapaz de seleccionar o identificar la interpretación correcta de la preferencia de W una vez que el traslado ha tenido lugar. Sin embargo, este autor sostiene que la clase de auto-conocimiento cuya infalibilidad garantiza el modelo basado en la inclusión de Burge está fuera del alcance de los argumentos incompatibilistas y sería suficiente, según él, para defender la plausibilidad del externismo psicológico (2003a, pp. 254-255).

acudiremos a los textos de Gibbons y Burge – (1996) y (1998), respectivamente – para examinar sus alegaciones contra la premisa (M3).

Autores que rechazan (M1)

Este subapartado estará dedicado al examen de las propuestas de Brueckner en (1997) y Ludlow (1995b) y (2004), cuyo objetivo común es el cuestionamiento de la primera premisa del argumento sobre la memoria.

En lo que respecta a las aportaciones del primero de estos autores, la tesis esencial de Brueckner consiste en que el argumento de Boghossian sólo es concluyente si añadimos ciertas cualificaciones adicionales en la primera premisa. De acuerdo con Brueckner, para que Óscar₁ sepa en t₂ qué es lo que pensaba en t₁ hace falta algo más que lo establecido en el antecedente de la premisa (M1), a saber, que no lo haya olvidado. La propuesta de este autor se apoya en la denominada “epistemología de las alternativas relevantes”, un análisis de la justificación epistémica que ha sido utilizado por algunos partidarios del compatibilismo.¹³⁹ En lo que sigue, vamos a explicar lo esencial de este análisis para ver de qué modo Brueckner se sirve de él en su argumentación.

Lo primero que debe señalarse en relación a la *epistemología de las alternativas relevantes* es que se trata de una teoría de carácter externista. Ya en el primer capítulo se dijo que el internismo y el externismo son denominaciones genéricas para agrupar una serie de teorías que pertenecen a distintos ámbitos de la filosofía. Así, de la misma manera que uno puede ser internista o externista en semántica o en psicología, también se puede adoptar uno de estos enfoques en el análisis de la justificación epistémica. En este último caso, lo disputado entre estas dos opciones es el carácter interno o en parte externo al agente de aquellos factores que hacen que una creencia verdadera sea, además, una creencia justificada. Según vimos, esta cuestión equivale a preguntar si los elementos que permiten considerar que la creencia de que P que mantiene un hablante S es una creencia justificada son *todos* ellos elementos a los que S tiene acceso epistémico. Mientras que un internista en epistemología daría una respuesta afirmativa a esta cuestión, los partidarios del externismo sostienen que algunos de esos elementos podrían situarse fuera del alcance epistémico de un S. En concreto, los externistas en

¹³⁹ Cf. Falvey y Owens (1994), Bernecker (1997) y Warfield (1997).

epistemología defienden que una creencia estará justificada si el hablante que la mantiene permanece en una relación externa (causal) apropiada con el objeto o estado de cosas acerca del cual versa su creencia. De este modo, el *externismo epistémico* o *fiabilismo* hace depender la justificación de una creencia de la fiabilidad del proceso externo que es causalmente responsable de ella, siendo la fiabilidad “la tendencia de un proceso a producir creencias que son verdaderas más bien que falsas” (Goldman 1976a, p. 345). Si los antecedentes causales de la creencia de que P que el sujeto S mantiene son tales que dan lugar a un número mayor de creencias verdaderas que de creencias falsas, entonces el proceso en el que dichos antecedentes están insertos es un proceso fiable y la creencia de que P será una creencia justificada.¹⁴⁰

El último paso antes de retomar la solución de Brueckner es explicar la noción de *alternativas relevantes* según ha sido utilizada por algunos partidarios del fiabilismo. Una alternativa relevante a una proposición P es, en primer lugar, una situación o estado de cosas Q tal que, de darse, haría que P fuera falsa. Si, por ejemplo, Clara señala a un perro y afirma “ese perro es un pastor alemán”, una alternativa relevante a la proposición expresada por esta preferencia sería una en la que, por ejemplo, Clara hubiera señalado a un mastín. Sin embargo, en un escenario en el cual un sujeto está en posición de conocer que P no todas las situaciones que, de darse, harían falsa esa proposición son alternativas relevantes, sino sólo aquellas que este sujeto sería capaz de discriminar con respecto al escenario actual. Dicho esto, podemos ayudarnos de una caracterización de las tesis fiabilistas que hace uso de esta noción de relevancia: la creencia de que P es una creencia justificada si y sólo si ha sido causada por un proceso fiable, en el sentido de que es un proceso que S puede distinguir de otras alternativas que harían que P sería falsa.¹⁴¹

¹⁴⁰ El fiabilismo cobra fuerza de una tesis considerablemente intuitiva sobre nuestras exigencias para decir que un sujeto conoce una determinada proposición, a saber, que si S mantiene la creencia verdadera de que P esta creencia no podría contar como conocimiento si la verdad de P es “cuestión de suerte”. Por ejemplo, si Iván afirma, sin haber realizado la más mínima comprobación al respecto, que hay cinco personas en su habitación y, por casualidad, acierta, no estamos dispuestos a decir que Iván *sabía* que hay cinco personas en su habitación. Podríamos pensar en una situación análoga. Supongamos que este individuo afirma “el objeto que está en aquella estantería es un valioso jarrón chino”. Sin embargo, él está situado en un entorno donde se han mezclado valiosos jarrones chinos junto con el mismo número de réplicas carentes de valor e indiscernibles de los primeros. Si casualmente este individuo acertara al señalar un jarrón chino, esto no sería suficiente para decir que *sabía* que *ese* era un jarrón chino. Lo interesante es apreciar que, en este último caso, el proceso por el cual nuestro individuo llega a sostener la creencia relevante no es un proceso fiable, puesto que al estar manipulado de la forma descrita tiende a producir tantas creencias verdaderas como falsas.

¹⁴¹ Veamos un sencillo ejemplo. Luis entra en una habitación y señala un bulto mientras profiere la oración “eso es un objeto azul”. Las condiciones lumínicas de ese entorno son pésimas y, sin embargo, nuestro individuo expresa una creencia verdadera porque el objeto señalado es, en efecto, de color azul.

Una vez dicho todo esto, retomamos el análisis de Brueckner destinado a neutralizar el argumento de Boghossian. Según dijimos, su propuesta consiste en impugnar el establecimiento de la premisa (M1) alegando que para que S sepa en t_2 lo que pensaba en t_1 no sólo se requiere que S no lo haya olvidado, sino que también se precisa el cumplimiento de un requisito adicional derivado del análisis fiabilista del conocimiento que hemos recogido en las líneas precedentes. Así, el traslado inadvertido de Óscar₁ a la Tierra Gemela encaja perfectamente en la noción de proceso no fiable de formación de creencias que el sujeto no puede discriminar como tal. En efecto, ambos mundos son, por definición, subjetivamente equivalentes para Óscar₁, ya que este sujeto es incapaz de distinguir entre un entorno y otro. Aplicando esto, supongamos que Óscar₁ es trasladado inadvertidamente de un planeta a otro en varias ocasiones. Si se le informara de esa situación y se le pidiera especificar qué concepto (agua o agua₂) estaba involucrado en la creencia que en un momento del pasado expresó al decir “recuerdo que creía que el agua es un constituyente fundamental de los seres vivos”, es muy probable que este sujeto no supiera qué decirnos, puesto que no ha advertido los cambios que se han producido en sus conceptos. Por esta razón, Brueckner sostiene que si Óscar₁ se decantara finalmente por el concepto de agua, no estaríamos dispuestos a decir que este hablante *sabía* que *ese* era el concepto que estaba expresando en su preferencia, aun cuando su aseveración fuera correcta. Como vemos, esta intuición se ve respaldada por el análisis fiabilista del conocimiento, ya que si S mantiene la creencia de que P, esta creencia sólo estaría justificada si S fuera capaz de excluir las alternativas relevantes a P, esto es, si pudiera discriminar la situación actual de otras situaciones alternativas en las que P es falsa. Pero, por principio, Óscar₁ carece de la habilidad para distinguir su situación actual, en la que está expresando el concepto de agua, de una situación alternativa en la que el concepto expresado fuese el concepto de agua₂.

Brueckner se sirve de estas observaciones para exigir el cumplimiento de una condición adicional como requisito necesario para que un sujeto sepa en t_2 qué pensamiento estaba pensado en t_1 . Este requisito, denominado *condición-m*, establece que ese hablante debe ser capaz de discriminar la situación en la que afirma que recuerda que P de otras situaciones alternativas en las que el pensamiento recordado

Según el análisis fiabilista, la creencia de Luis no estará justificada, puesto que el proceso de generación de la misma no es fiable: este individuo es incapaz de distinguir esa situación de una situación alternativa en la que el bulto señalado es de color rojo.

tiene un contenido distinto. De esta manera, la premisa que debería figurar en el argumento de Boghossian en lugar de (M1) debería ser (M1)':

(M1)' Si en t_2 S no ha olvidado nada de lo que sabía en t_1 , y si S satisface la condición-m para el conocimiento basado en la memoria de aquellas proposiciones que conocía en t_1 , entonces lo que S sabe en t_1 también lo sabe en t_2 (Brueckner 1997, p. 330).

Ahora bien, como la situación epistémica de Óscar₁ no satisface la condición-m para el conocimiento basado en sus recuerdos, el establecimiento de la segunda premisa (M2), que afirma que este sujeto no ha olvidado nada, es insuficiente para extraer la conclusión incompatibilista.

Una vez analizada la réplica de Brueckner, examinaremos las sugerencias que Ludlow nos presenta en (1995b) y en (2004) con el objeto de demostrar que el argumento de Boghossian no es concluyente. Para empezar, diremos que existe un estrecho parentesco entre las propuestas que Ludlow ofrece en ambos trabajos, en la medida en que la aportación principal de (1995b) sigue vigente en su artículo de (2004), aun cuando este artículo presenta una solución un tanto más sofisticada. Dicha aportación consiste en una auto-denominada “explicación externista de la memoria” cuyo fin es impedir el establecimiento de la primera premisa del argumento incompatibilista de Boghossian, a saber, que si S no olvida nada, lo que sabe en t_1 debe saberlo en t_2 . Podemos enunciar el contenido de esta explicación de la siguiente manera:

(EEM) Las condiciones de identidad (externas en parte) de los contenidos que figuran en los actos de rememoración son aquellas que pertenecen al entorno en que se lleva a cabo dicho acto.

Veamos cómo se aplica este principio al caso de traslado que estamos examinando. Cuando el protagonista del mismo, Óscar₁, se encontraba en la Tierra en el tiempo t_1 , expresaba una creencia sobre el agua al proferir “el agua es un símbolo de purificación”. En un momento dado, este agente es víctima de un traslado inadvertido a un nuevo entorno, la Tierra Gemela, donde *ya no hay agua*. Como el factor externo que determinaba la identidad de esa creencia ya no está presente y en su lugar figura un factor distinto, entonces la creencia que Óscar₁ expresará a partir de ahora en la Tierra Gemela al decir “el agua es un símbolo de purificación” será una creencia diferente. Supongamos, además, que este sujeto emite en t_2 – tiempo en el cual se sitúa en la Tierra Gemela – la siguiente preferencia “en la escuela me enseñaron que el agua es un símbolo de purificación”. Al decir esto, Óscar₁ pretende estar rememorando un

pensamiento sobre el agua, puesto que remite al líquido terrestre. Lo que la tesis (EEM) enuncia es que las características naturales de la Tierra Gemela son responsables de la individuación de *todos* los pensamientos que Óscar₁ expresa haciendo uso del término “agua” en ese entorno, incluidos aquellos pensamientos que adquirió en un entorno diferente y cuyos contenidos, por tanto, fueron individuados originariamente a partir de factores naturales distintos. Con otras palabras, dado que en la Tierra Gemela no hay agua, los pensamientos que Óscar₁ expresará haciendo uso del término “agua” en este nuevo entorno no podrán ser pensamientos sobre ese líquido, aun cuando se trate de pensamientos rememorados que versaban acerca del agua en el momento de su adquisición.

La propuesta de Ludlow consiste en aplicar la tesis (EEM) junto con el modelo basado en la inclusión para impedir el establecimiento de la primera premisa del argumento de Boghossian. Según Ludlow, Óscar₁ siempre sabe lo que piensa en el momento de pensarlo. Las condiciones externas de la Tierra determinan que la creencia que Óscar₁ expresa en t_1 mediante la preferencia “el agua es un símbolo de purificación” sea una creencia sobre el agua. Además, *esas mismas condiciones externas* son las que fijan la identidad del contenido del pensamiento de segundo orden que Óscar₁ se forma al reflexionar sobre ella y tomarla como contenido en ese entorno. En t_2 , sin embargo, las condiciones externas han cambiado. Al estar inserto en la Tierra Gemela, la creencia que Óscar₁ expresa al hacer uso del término “agua” será una creencia distinta y, según el modelo basado en la inclusión, también lo serán los pensamientos reflexivos que Óscar₁ mantiene sobre la misma.

En síntesis, el núcleo de la propuesta de Ludlow consiste en afirmar que las condiciones externas del nuevo entorno fijan el contenido de los pensamientos rememorados, aun cuando se trate de pensamientos que fueron adquiridos en otro entorno y que tuvieron un contenido distinto en el momento de su adquisición. Podemos describir esto esquemáticamente:

Tiempo	Entorno	Preferencia	Pensamiento expresado
T ₁	Tierra	“El agua es un símbolo de purificación”	Creencia acerca del agua
T ₂	Tierra Gemela	“Recuerdo que en la escuela me enseñaron que <i>el agua es un símbolo de purificación</i> ”	Recuerdo de una creencia acerca del agua ₂

¿En qué medida estas consideraciones impiden la conclusión del argumento de la memoria? Dada la tesis (EEM), el cambio de entorno conlleva que el pensamiento que Óscar₁ adquiere en t₁ es un pensamiento distinto del pensamiento que trae a su memoria en t₂. Pero, si esto es así, entonces este hablante no sabe en t₂ lo que pensaba en t₁. Dado que Ludlow acepta la premisa (M3) del argumento de Boghossian y cree, por tanto, que Óscar₁ no ha olvidado nada entre t₁ y t₂, entonces podemos establecer la siguiente conjunción:

(L) Óscar₁ no ha olvidado nada y tampoco sabe en t₂ lo que pensaba en t₁.

Como puede apreciarse, (L) impide el establecimiento de (M1), la premisa que afirma que si un hablante no ha olvidado nada, lo que sabe en t₁ debe saberlo en t₂.¹⁴²

Sin embargo, el propio Ludlow reconoció en (2004) su insatisfacción con la solución ofrecida en (1995b), no tanto porque crea incorrecta su estrategia anterior sino porque considera que no ha ido suficientemente lejos. En este sentido, este autor defiende que sólo podremos evitar las conclusiones de Boghossian si *además* de aceptar (EEM), como nos pedía en (1995b), pensamos de una manera diferente la relación entre pensamientos y contenidos, una relación que, típicamente, asigna a cada pensamiento un único contenido, el cual, a su vez, determina la identidad del pensamiento. El autor sugiere que sería mejor describir nuestros pensamientos como pensamientos-matriz, es decir, como pensamientos que albergan, cada uno de ellos, una multiplicidad de contenidos indexados en el tiempo (Ludlow 2004, p. 424).

El mejor modo de introducir la nueva propuesta de Ludlow es analizar las tensiones que presentaba su solución de (1995b). El conflicto reside en la conjunción de estas tres tesis:

¹⁴² Hay que decir, sin embargo, que la solución de Ludlow no satisfizo a demasiados compatibilistas. Brueckner afirma en (1997) que existe una diferencia entre lo que Ludlow dice estar haciendo en (1995b) y lo que hace de hecho en ese trabajo. Así, este autor *afirma* que la premisa débil de la argumentación de Boghossian es (M1) y, como hemos visto, su estrategia para impedir el establecimiento de esta premisa pasa por demostrar la plausibilidad de (L), una tesis incompatible con (M1). Ahora bien, Brueckner sostiene que lo que *en realidad* se desprende de las consideraciones de Ludlow es la falsedad de la premisa (M2). La razón es que la tesis (EEM) implica que los pensamientos rememorados en t₂ son distintos de los pensamientos adquiridos en t₁ y, como Óscar₁ no recuerda qué es lo que pensaba en el pasado (porque trae a su memoria pensamientos distintos), entonces su memoria falla. En consecuencia, el protagonista del caso de traslado ha olvidado lo que pensaba en t₁, contrariamente a lo que sostiene (M2). No obstante, cabe hacer una crítica de otro signo. A nuestro modo de ver, la motivación principal para rechazar la solución de Ludlow consiste en que la explicación externista de la memoria recogida en (EEM) impide que Óscar₁ pueda recordar en t₂ los pensamientos verdaderos sobre el agua que mantuvo en t₁ y, por tanto, dicha explicación legitimará un fallo sistemático de la memoria de este individuo (Ludlow 1996, p. 314), algo muy poco deseable.

(TL1) Un pensamiento posee un único contenido que determina su identidad.

(TL2) Óscar, el protagonista del caso de traslado, no ha olvidado nada.

(EEM) Las condiciones de identidad de los contenidos que figuran en los actos de rememoración son aquellas que pertenecen al entorno en que se lleva a cabo dicho acto.

La dificultad para compatibilizar todos estos elementos parece clara. La tesis (EEM) conlleva que la identidad de los contenidos mentales que un sujeto S rememora viene dada por las condiciones del entorno en el que se lleva a cabo la rememoración. Esto supone que S podría traer a su memoria un contenido distinto del que pretende estar recordando si se dieran diferencias relevantes entre el entorno en el que adquirió el contenido y el entorno en el que lo recuerda. Además, según (TL1), estas diferencias en el contenido suponen que los pensamientos que S mantiene en cada entorno son distintos, es decir, que todos aquellos recuerdos que expresa haciendo uso del término “agua” serían falsos. Pero, de aceptar esto, Óscar₁ no habría conseguido recordar de manera fidedigna qué pensó en el pasado. Lo paradójico es que, a pesar de su incapacidad para traer a la memoria su pensamiento pasado, Ludlow sostiene que Óscar₁ no ha olvidado nada, contrariamente a lo afirmado en (TL2).

Ante esta situación tan incómoda, este autor opta por renunciar a (TL1), uno de los supuestos fundamentales de las discusiones sobre la identidad metafísica de nuestros pensamientos.¹⁴³ Lo que nos propone es entender cada uno de nuestros pensamientos como una *matriz de varios contenidos*. De esta manera, si Óscar₁ rememora en t_2 un contenido distinto del adquirido en t_1 – debido a la diferencia del entorno en que está inserto – esto no impide que sepa lo que piensa, es decir, que pueda tener en ambos casos el mismo pensamiento. Para ello, bastaría con que el contenido adquirido y el contenido rememorado fueran contenidos pertenecientes a un mismo pensamiento-matriz. Éste es el modo en que el autor consigue preservar (TL2) y afirmar, con ello, que Óscar₁ no ha olvidado nada entre un entorno y otro, algo que desde las coordenadas de (1995b) era muy difícil mantener. La solución de Ludlow, en suma, pretende conjugar el externismo y el acceso privilegiado a los propios pensamientos por la vía de permitir que dicho acceso tome en consideración en cada momento *sólo uno* de los distintos contenidos involucrados en el pensamiento. Además, Ludlow parece encontrar una justificación independiente para defender su visión de los pensamientos en tanto

¹⁴³ Ludlow es consciente de la radicalidad de su tesis y admite que tiene algo de “especulativa” (2004, p. 421).

que matrices de varios contenidos. Según este autor, la finalidad de las atribuciones de contenidos es hacer inteligible el comportamiento de un agente S para un espectador H (Ludlow 2004, p. 421) y si dos espectadores atribuyen a S dos contenidos distintos que, sin embargo, explican igualmente bien las acciones que S lleva a cabo, entonces se tratará de dos atribuciones legítimas porque nos permiten capturar (cada una mediante un contenido distinto) el pensamiento de S.

En nuestra opinión, la solución de Ludlow es insatisfactoria y esto es debido a que sólo ha conseguido neutralizar la conclusión incompatibilista de Boghossian al precio de renunciar al externismo. Desde luego, esta conclusión no es algo que Ludlow admita en (2004), ya que este autor considera que (EEM), tesis que sigue asumiendo en este trabajo, es una explicación *externista* de la memoria. Por ello, de ser correctas nuestras consideraciones, su solución debería ser objeto de revisión.

Para justificar esta observación vamos a aclarar un punto de la propuesta de Ludlow acerca del cual él mismo no proporciona demasiadas indicaciones. Lo que buscamos es averiguar cuál es el *denominador común* de una serie de contenidos (c_1 , c_2 , c_3 , etc.) que pertenecen a un mismo pensamiento-matriz P o, lo que es igual, cuál es el criterio de identidad de esta clase de pensamientos. Así, resulta natural pensar que lo mismo que determina que los contenidos c_1 , c_2 y c_3 sean contenidos de un mismo pensamiento será lo que nos permita explicar por qué las variaciones entre estos contenidos no modifican la identidad del pensamiento-matriz que los contiene, puesto que tales contenidos c_1 , c_2 y c_3 resultarán igualmente apropiados para caracterizar el pensamiento P de un sujeto S. Creemos que la mejor manera de averiguar esto consiste en retomar la clase de problemas que Ludlow pretende resolver con su propuesta de los pensamientos-matriz. Tomemos en consideración un ejemplo parecido al que nos propone este autor en (2004). Ludlow parte de la preferencia que Óscar₁ emite en la Tierra Gemela, una vez que se produce el traslado:

(O) El agua es un líquido a temperatura ambiente.

Según este autor, un hablante de la Tierra Gemela, Juan₂, interpretará el pensamiento que Óscar₁ expresa mediante (O) como un pensamiento acerca del agua₂. Si un hablante de la Tierra, Juan₁, hubiera escuchado en ese momento la preferencia (O) de labios de Óscar₁, habría atribuido a Óscar₁ un pensamiento con un contenido que versa sobre el agua. Estos serían los informes que un hablante y otro elaborarían para caracterizar el pensamiento de Óscar₁:

Juan₁: “Óscar₁ cree que el agua es un líquido a temperatura ambiente” [siendo el contenido de esta atribución que Óscar₁ cree *que el agua (H₂O) es un líquido a temperatura ambiente*].

Juan₂: “Óscar₁ cree que el agua es un líquido a temperatura ambiente” [siendo el contenido de esta atribución que Óscar₁ cree *que el agua₂ (XYZ) es un líquido a temperatura ambiente*].

Lo interesante es darnos cuenta de que aun cuando estos informes atribuyen contenidos diferentes a Óscar₁ ambos resultan igualmente apropiados, según Ludlow, para caracterizar el pensamiento que expresa mediante la preferencia de (O). Esto es así porque ambos contenidos pertenecen a mismo pensamiento-matriz.

Su solución a los casos de traslado puede proporcionarnos otras pistas para responder a la pregunta por el denominador común de los contenidos que pertenecen a una misma matriz. Este autor pretendía esquivar la conclusión incompatibilista de estos argumentos proponiendo que los contenidos que figuran en el pensamiento adquirido en t_1 y en el pensamiento recordado en t_2 son contenidos que, en el fondo, pertenecen a un mismo pensamiento. Por tanto, sea cual sea el contenido que Óscar₁ se auto-atribuya en esos dos instantes del tiempo – un contenido sobre el agua o un contenido sobre el agua₂ – el protagonista del caso de traslado conservará en todo momento su acceso epistémico al pensamiento en cuestión. Lo que tienen en común los contenidos atribuidos por Juan₁ y Juan₂, por una parte, y los contenidos que Óscar₁ se auto-atribuye en distintos momentos del tiempo es, precisamente, que ambos son contenidos que versan sobre el agua o sobre el agua₂, esto es, sobre entidades que le resultan a Óscar₁ *completamente indistinguibles desde su punto de vista*. A este respecto conviene recordar que una de las premisas de los casos de traslado consiste en que Óscar₁ no es capaz de advertir que en su nuevo entorno su preferencia “el agua es un líquido a temperatura ambiente” ha pasado a expresar un contenido diferente del que expresaba en la Tierra, y esto sólo se explica si el nuevo contenido es, por lo que a él concierne, completamente indistinguible del contenido que expresaba en la Tierra mediante el uso de la misma expresión. Dicho de otra manera, Óscar₁ debe encontrarse en t_2 en el mismo estado interno, de carácter cualitativo o fenomenológico que se encontraba en t_1 y esta identidad tiene su origen en el hecho de que, *ex hypothesi*, el agua y el agua₂ son

líquidos que comparten todas sus propiedades superficiales y cuyo factor diferenciador le resulta desconocido a Óscar₁.¹⁴⁴

Podemos enunciar explícitamente el criterio de identidad para los pensamientos-matriz del siguiente modo. Consideraremos que dos hechos h y h' son contrapartidas epistémicas para un sujeto S en un tiempo t si y sólo si sitúan a S en el mismo estado cualitativo en t (Sawyer 1999, p. 361). Tanto el agua₁ como el agua₂ son, según esta definición, contrapartidas epistémicas del agua para Óscar₁. A partir de esto, podemos decir que un pensamiento-matriz P de un sujeto S que tenga como uno de sus contenidos c_1 y remita al estado de cosas e_1 albergará además (y solamente) aquellos contenidos ($c_2, c_3, \text{etc.}$) que remitan a estados de cosas ($e_2, e_3, \text{etc.}$) que sean para S en t contrapartidas epistémicas de e_1 .

Ahora bien, sostenemos que este criterio de identidad para los pensamientos no es un criterio genuinamente externista. La mejor manera de demostrarlo es confrontar la idea del pensamiento-matriz con las bases de un experimento mental tipo a favor de la individuación externista de nuestros estados mentales. Si esa clase de experimentos permiten reforzar la plausibilidad del propio externismo es porque nos instan a concluir que dos sujetos que comparten todos sus aspectos internos mantienen, sin embargo, pensamientos diferentes en función de las diferencias (naturales o sociales) que existen entre los entornos en los que uno y otro están situados. Es decir, si pueden existir diferencias psicológicas entre dos hablantes internamente idénticos, entonces, concluye el externista, los factores internos no agotan la determinación de nuestros contenidos mentales. Sin embargo, la propuesta de Ludlow consiste en afirmar que si uno de los hablantes es trasladado al entorno en el que habita su réplica – un mundo fenomenológicamente idéntico a su entorno de origen –, sus contenidos cambiarán con

¹⁴⁴ Para confirmar esta idea, podríamos ver cómo funciona la estrategia de Ludlow en una prolongación del caso de traslado en el que se transportara sin su conocimiento a Óscar₁ a un tercer entorno, la Tierra Gemela-Bis, donde todo es idéntico a la Tierra excepto un aspecto, a saber, que el líquido inodoro, incoloro e insípido que corre por sus ríos y lagos y sacia la sed de sus habitantes, líquido que todo los habitantes de este nuevo entorno denominan “agua”, es, en realidad, esencialmente distinto al agua y tiene la composición química ABC. En esta situación, si un habitante de la Tierra Gemela-Bis, José₃, dijera de Óscar₁ “él cree que el agua es un líquido a temperatura ambiente”, atribuyéndole con ello un contenido – relativo al agua₃ – distinto del que le atribuían Juan₁ y Juan₂, esta atribución sería, según la estrategia de Ludlow, igualmente apropiada para caracterizar el pensamiento de Óscar₁. Cuando este hablante llevara a cabo un acto de rememoración en t_3 mientras emite una preferencia como “recuerdo que en la escuela me enseñaron que el agua es un símbolo de purificación”, las pautas de Ludlow nos permiten establecer que, en virtud de (EEM), el contenido rememorado versará sobre el agua₃ y que, además, seguiría tratándose de un contenido legítimo para capturar su pensamiento. Lo que vemos, en consecuencia, es que todos estos contenidos que hemos mencionado (sobre el agua, sobre el agua₂, sobre el agua₃) serán contenidos que nos sirven para capturar el pensamiento P de Óscar₁ porque se trata de contenidos idénticos *por lo que a él respecta*.

el cambio de su entorno y, *sin embargo, la identidad de sus pensamientos permanecerá inalterada*. Pero, si esto es así, entonces no existe una verdadera diferencia entre los contenidos psicológicos de los pensamientos de este sujeto y los contenidos de los pensamientos de todas sus réplicas internas contrafácticas, por lo que las diferencias en los factores externos no serían psicológicamente significativas.¹⁴⁵

En resumen, la sugerencia de Ludlow, que consiste en permitir la atribución de distintos contenidos a un agente que pueden servir para caracterizar el mismo pensamiento P, tiene consecuencias poco favorables para el externismo. Si de lo que se trata es de encontrar un rango de variación admisible entre los contenidos c_1 , c_2 y c_3 que no afecte a la identidad de P y Ludlow lo sitúa en la compatibilidad de todos estos contenidos con la historia fenomenológica o interna de dicho agente, entonces se desprenderá que las variaciones en el entorno externo no suponen en ningún caso una variación en la identidad de su pensamiento P, a menos que estas variaciones afecten a la historia fenomenológica interna de dicho sujeto. Como puede apreciarse, esta conclusión tiene un carácter fuertemente internista.

Autores que rechazan (M2)

En este apartado examinaremos la solución que M. Tye propone en (1998) para esquivar el argumento de Boghossian, una solución que no ha recibido demasiado respaldo por parte de los autores compatibilistas. Esta propuesta consiste básicamente en asumir la tesis (EEM) que Ludlow propone en (1995b) pero admitiendo que se ha producido un fallo en la memoria de Óscar₁. De ser así, la premisa (M2) del argumento de Boghossian, que afirma que Óscar₁ no olvida nada entre t_1 y t_2 , habrá de ser rechazada y, consiguientemente, también se rechazará la conclusión de dicho argumento.

Las razones que Tye ofrece para reforzar la explicación externista de la memoria propuesta por Ludlow son dos. En primer lugar, afirma que sólo podremos dar sentido a ciertos *juicios comparativos* de Óscar₁ si permitimos que los factores externos del entorno presente sean los que determinen el contenido de los pensamientos rememorados. Tye nos propone el siguiente ejemplo. Supongamos que el protagonista

¹⁴⁵ Esta consecuencia puede extenderse al resto de los experimentos mentales que suelen emplearse a favor del externismo. De este modo, dado que Bert y Bert-contrafáctico – los protagonistas del experimento mental de Burge – son réplicas exactas en sus aspectos fenomenológicos, la aplicación de las tesis de Ludlow sobre los pensamientos-matriz nos lleva a establecer una identidad en el contenido de los pensamientos de ambos individuos.

de los casos de traslado emite la siguiente preferencia en la Tierra Gemela, cuando el traslado ha tenido lugar:

(T) “Lo único que bebo después de las cinco de la tarde es agua. Sin embargo, hace años solía beber agua reforzada con ginebra por las tardes. Me lo pasaba muy bien aquellas tardes. El agua mejora mucho al mezclarla con ginebra”.

Al emitir esta preferencia, Óscar₁ pretende establecer un contraste entre sus hábitos pasados y presentes. Dicho contraste, nos dice Tye, sólo tiene sentido bajo la presuposición de que el término “agua” es unívoco en (T), es decir, si el concepto que expresa el término “agua” en todas sus apariciones es uno y el mismo. Dado que la aplicación de la tesis (EEM) a (T) tiene como consecuencia que el término “agua” expresa el mismo concepto (agua₂) en sus dos apariciones y nos permite respetar con ello la pretensión comparativa de Óscar₁, inferimos (EEM) como la mejor explicación de la individuación de los contenidos de este hablante.

El segundo argumento que proporciona Tye a favor de (EEM) apela al anti-individualismo de Burge. De este modo, como los pensamientos que mantiene un sujeto dependen, en parte, del uso que su comunidad lingüística determina para los términos que figuran en la expresión de sus pensamientos, hay una razón adicional para considerar que los pensamientos que Óscar₁ expresa *ahora* en la Tierra Gemela son pensamientos acerca del agua₂ y es que la comunidad lingüística de ese entorno acuerda unas pautas de utilización del término “agua” distintas de las terrestres. En concreto, en el entorno de acogida de Óscar₁ se establece que ese término debe designar el agua₂. Si esto es así, el concepto que debe figurar en los pensamientos que este hablante expresa mediante el término “agua” es el concepto de agua₂, incluso aunque estos pensamientos remitan a pensamientos rememorados que originariamente contenían el concepto de agua, tal y como recoge (EEM).

Por otro lado, ya hemos dicho que Tye sostiene, al contrario que Ludlow, que la tesis (EEM) implica que se ha producido un fallo en la memoria de Óscar₁. Así, este autor alega que el cambio entre entornos hace que la memoria se vuelva “defectuosa”, en el sentido de que las circunstancias del entorno externo *cambian* el contenido de los recuerdos (Tye 1998, p. 90). A este respecto él trata de reforzar la plausibilidad de esta conclusión afirmando que la función de la memoria no es “fotografiar” pensamientos pasados para mantenerlos “congelados” en el recuerdo.¹⁴⁶ Como la memoria no es una

¹⁴⁶ Una observación similar puede leerse en Ludlow (1996), p. 316.

capacidad que funcione a partir de materiales puramente fenomenológicos, sino que involucra esencialmente conceptualización, de darse una variación en los conceptos responsables de configurar los recuerdos, es razonable suponer que los contenidos recordados también habrán cambiado.

No obstante, existe un serio problema con la estrategia de Tye, que explica por qué su solución a los casos de traslado no ha conseguido recabar demasiados apoyos. Este problema consiste en que el mal funcionamiento de la memoria, que Tye acepta como una consecuencia de la tesis (EEM), no es algo deseable en absoluto, aunque él intente desacreditar esta objeción al situar su origen en una concepción simplista de la memoria (como facultad de “fotografiar” pensamientos). En nuestra opinión no es conveniente legitimar un trastorno sistemático de los contenidos rememorados de Óscar₁, e impedir, con ello, que el protagonista de los casos de traslado recuerde todos aquellos pensamientos acerca del agua que adquirió en la Tierra. Esta réplica no se origina en una concepción “fotográfica” de la memoria, sino en la exigencia mínima de dotar a esta facultad de la capacidad para traer a la mente pensamientos pasados sin que esas reflexiones sean indefectiblemente falsas.

Autores que rechazan (M3)

En este apartado analizaremos algunas de las réplicas compatibilistas más importantes al argumento de Boghossian que han optado por cuestionar la premisa (M3). En concreto, examinaremos las propuestas de Burge en (1998) y Gibbons en (1996), quienes se oponen frontalmente a la explicación externista de la memoria ofrecida por Ludlow. En particular, ambos autores se niegan a aceptar que si se diera una variación entre el entorno en que se adquiere un pensamiento y el entorno en el que se rememora posteriormente, deban ser las características del entorno en el que se lleva a cabo la rememoración las condiciones *sistemáticamente* responsables de la individuación del contenido del pensamiento rememorado. Las consecuencias derivadas de mantener lo contrario son, como vimos, un tanto contraintuitivas, en la medida en que debería admitirse un fallo automático de la memoria en todas las reflexiones sobre los pensamientos pasados relevantes. Sin embargo, tanto Burge como Gibbons no consideran aceptable esta clase de fallo.¹⁴⁷

¹⁴⁷ Debemos señalar que la insatisfacción de Burge con una solución basada en la tesis (EEM) de Ludlow no tiene tanto que ver con que Burge rechace que en los recuerdos terrestres de Óscar₁ que rememora en

Un buen recurso para abordar el análisis de estos dos autores lo constituye la distinción que propone Goldberg en su trabajo de (2003a) entre dos tipos de explicaciones de los casos de traslado, a saber, explicaciones de *sustitución o reemplazamiento conceptual* y explicaciones de *suplementación conceptual*.¹⁴⁸ Al primer tipo de explicaciones pertenecen las tesis de aquellos autores que, después de confrontar el externismo con la situación de traslado inadvertido de Óscar₁, concluyen que dicho traslado conlleva la pérdida del concepto terrestre de agua y la adquisición de un concepto distinto que reemplaza al anterior, a saber, el concepto de agua₂. Tanto Ludlow como Tye son autores que se adhieren a la explicación de sustitución conceptual. Sin embargo, cabe sostener que el cambio en el repertorio conceptual resultante del traslado supone la adquisición de un nuevo concepto sin pérdida de los anteriores. Esto es lo que propone la explicación por suplementación conceptual, de la cual Gibbons es un firme partidario y Burge lo es en la mayoría de los casos. Este último afirma, concretamente, que el traslado puede tener distintos efectos en el repertorio conceptual de los sujetos que lo protagonizan. Podría ocurrir, por ejemplo, que el transporte implicara la formación de un concepto de *tipo disjunto*, esto es, que Óscar₁ pasara a expresar dos conceptos distintos con extensiones disjuntas haciendo uso del mismo término. Pero también existe la posibilidad de que el concepto que Óscar₁ expresa mediante el término “agua” se convirtiera, tras su viaje, en un *concepto amalgama*, esto es, un único concepto que se aplica tanto al H₂O como al XYZ (Burge 1998, p. 352).¹⁴⁹

Una vez admitido que el traslado no tiene por qué suponer *siempre* una pérdida conceptual, parece más fácil restar plausibilidad a la tercera premisa del argumento de Boghossian, que afirma que Óscar₁ no sabe que P en t₂. Así, según Gibbons, “si viajar a la Tierra Gemela no me priva de aquellos conceptos que tuve una vez, no hay razón para pensar que la creencia disposicional de segundo orden cambie en su contenido” (1996, p. 303). En efecto, tanto Burge como Gibbons consideran que en el caso concreto de Óscar₁ este hablante puede recordar sin problemas en t₂ el pensamiento

la Tierra Gemela pueda figurar el concepto de agua₂, como con el hecho de que la tesis (EEM) afirme que en *todos* los casos sería ese el concepto que figuraría en los recuerdos de Óscar₁. Es decir, Burge admite que existe la posibilidad (marginal, eso sí) de que un individuo no vuelva a activar sus recuerdos terrestres y, por tanto, pierda la habilidad de pensar específicamente en el agua. En tal caso, la premisa del argumento de Boghossian que deberíamos rechazar sería (M2), esto es, la premisa que afirma que Óscar₁ no olvida nada entre t₁ y t₂. Cf. Burge (1998), p. 369.

¹⁴⁸ Cf. Goldberg (2003a), pp. 245-246.

¹⁴⁹ Gibbons, por su parte, no distingue explícitamente entre tipos disyuntivos y tipos amalgama, como hace Burge (Gibbons 1996, p. 299).

que adquirió en t_1 , al conservar en su mente los conceptos terrestres relevantes.¹⁵⁰ Y, lo que es más importante, ambos defienden que Óscar₁ *sabe* en t_2 qué pensaba en t_1 apelando al modelo de la inclusión reflexiva.

Con objeto de examinar las estrategias de ambos autores, tomaremos en consideración primero las propuestas de Gibbons. Este autor sostiene que existe una *historia causal* que vincula el pensamiento adquirido en t_1 con el acto de rememoración de este pensamiento que Óscar₁ lleva a cabo en t_2 (Gibbons 1996, p. 292). Además, de acuerdo con el modelo basado en la inclusión, como el pensamiento que figura en esa reflexión es, precisamente, el pensamiento que Óscar₁ adquirió en t_1 , su propia reflexión heredará los mismos conceptos que figuran en este pensamiento de primer orden. Podemos poner en claro esta propuesta utilizando otra tabla:

Tiempo	Entorno	Proferencia	Pensamiento expresado
T ₁	Tierra	“El agua es un símbolo de purificación”	Creencia acerca del agua
T ₂	Tierra Gemela	“Recuerdo que en la escuela me enseñaron que <i>el agua es un símbolo de purificación</i> ”	Recuerdo de una creencia acerca del agua

De acuerdo con esto, Óscar₁ puede recordar en t_2 el *mismo* pensamiento que adquirió en t_1 y, en consecuencia, será capaz de saber qué es lo que pensaba en ese momento. Por ello, si Óscar₁ sabe en t_2 qué pensaba en la Tierra, entonces no podremos establecer la premisa (M3) del argumento de Boghossian y no habrá lugar para la conclusión incompatible.

En el caso de Burge, el vínculo entre el pensamiento adquirido en t_1 y el acto de rememoración que Óscar₁ lleva a cabo en t_2 se explica por la existencia de un cierto tipo de memoria, a saber, la *memoria preservativa* (Burge 1998, p. 357). Su función es “retener” los pensamientos adquiridos en el pasado estableciendo *cadena causales memorísticas* entre estos pensamientos y los actos futuros de rememoración de los mismos. Eso sí, Burge advierte de que esa clase de memoria no sirve para “identificar” o “referirse” a los contenidos, pensamientos o imágenes adquiridos en el pasado. Tal matización tiene que ver con el hecho de que este autor, como señalamos, admite que el cambio conceptual derivado del traslado es un cambio que le resulta inadvertido a

¹⁵⁰ Burge sostiene, como vimos, que la formación de un concepto de tipo disjunto y el olvido del concepto anterior se produce cuando un individuo no reactiva los recuerdos de su entorno original. Esta clase de situaciones no son la norma y no hay motivos para contemplar de esa manera la situación de Óscar₁.

Óscar₁. Por ello, aunque se informara a este agente de que el traslado ha tenido lugar, sería incapaz, a pesar de ello, de señalar a partir de qué momento el concepto de agua₂ es el que figura en las creencias que expresa haciendo uso del término “agua”. Ahora bien, esto no significa, según Burge, que Óscar₁ no sabe ahora qué pensamientos mantenía en el pasado en el sentido que le interesa a Boghossian. Para Burge la memoria preservativa es lo suficientemente potente como para retener el pensamiento pasado sin alterar su identidad, aunque no permita a Óscar₁ identificar ese pensamiento pasado como un pensamiento distinto del que expresa en la actualidad al proferir “el agua es un símbolo de purificación”.

En definitiva, estos autores han encontrado un modo de esquivar el argumento incompatibilista de Boghossian mucho menos costoso que la explicación externista de la memoria propuesta por Ludlow y adoptada por Tye. Mientras que estos últimos aceptaban que Óscar₁ había perdido *todos* sus recuerdos sobre el agua una vez que ha tenido lugar el traslado y el consiguiente cambio conceptual, Burge y Gibbons se aseguran un buen funcionamiento de la memoria al tiempo que soslayan la conclusión escéptica del argumento de Boghossian.

Antes de comenzar la segunda sección de este capítulo, ofreceremos una breve recapitulación de los aspectos más relevantes de la discusión tratada hasta el momento. Las respuestas que los partidarios del compatibilismo han ofrecido a los razonamientos de Boghossian de (1989) son bastante diferentes entre sí, aunque pueden detectarse algunas constantes que nos proporcionarán una pista sobre los aspectos que serán destacados en las siguientes fases del debate. Como hemos indicado, Brueckner mantenía que la situación del hablante que, tras el traslado, pretende recordar sus pensamientos terrestres, es una situación que incumple las condiciones que harían que esa rememoración fuera un proceso epistémico fiable. Por otra parte, las réplicas de Ludlow y Tye, por un lado, y de Burge y Gibbons, por otro, se encuentran en abierta contradicción con respecto al papel del entorno externo actual en la individuación de los contenidos rememorados. Según señalamos, para los dos primeros estas condiciones externas son las responsables de la individuación de todos los contenidos de la víctima de traslado, incluidos los contenidos rememorados que fueron adquiridos en un entorno externo que difiere de manera relevante del entorno externo de la rememoración. En cambio, Burge y Gibbons argumentan que existen cadenas causales memorísticas que permiten que esta última clase de contenidos preserve su identidad. Burge, en particular, habla de la “memoria preservativa” como la facultad encargada de traer al presente

aquellos contenidos adquiridos en el pasado sin necesidad de referirse a ellos, esto es, de determinar si son iguales o distintos de los contenidos presentes. La solución de Burge y Gibbons es, a nuestro modo de ver, una propuesta mucho más plausible que la “explicación externista de la memoria”. El motivo es que esta clase de explicación impedía que Óscar₁ tuviera acceso privilegiado a sus contenidos terrestres que involucraban el concepto de agua y, dado que el objetivo de una argumentación compatibilista es conservar intacto en la medida de lo posible el acceso privilegiado dado el externismo, podemos concluir que la propuesta de Ludlow no es apta para lograr este objetivo. Por otro lado, a pesar de que su argumentación de (2004) se presentaba a sí misma como radical y programática, escondía en realidad un movimiento extremadamente defensivo, hasta el punto de hacer psicológicamente irrelevantes aquellas diferencias en el contenido mental que no alteraran la historia interna de los sujetos, algo que, según hemos argüido, suponía caer de lleno en el internismo.

Dicho esto, podemos destacar, sin embargo, la existencia de una cierta coincidencia en las distintas respuestas dirigidas a Boghossian. Varios autores compatibilistas reconocen explícitamente que el externismo no permite conservar la capacidad de un hablante para discriminar entre sus contenidos y averiguar si se trata de los mismos o no, aunque todos ellos se esfuerzan en justificar que esta carencia es inocua y no implica, *per se*, que el externismo sea incompatible con el acceso privilegiado.¹⁵¹ El núcleo de las discusiones posteriores sobre los casos de traslado será, precisamente, si la falta de habilidades discriminativas en el auto-conocimiento es tan inofensiva como estos autores compatibilistas declaran.

2. *Casos de traslado y razonamiento*

Vamos a dedicar esta sección a examinar el argumento incompatibilista que Boghossian plantea en (1992) y (1994), así como las réplicas formuladas por algunos externistas para contrarrestar sus conclusiones. El objetivo de estos nuevos trabajos de Boghossian es ofrecer una nueva reducción al absurdo de las tesis externistas sin entrar en discusiones sobre el funcionamiento de la memoria. Esta vez el autor intentará mostrar

¹⁵¹ Cf. Burge (2003a), p. 282; Falvey y Owens (1994), p. 112; Goldberg (2003a), pp. 254-255 y Bernecker (1997), p. 343.

que la falta de habilidades discriminativas en Óscar₁ hace que este sujeto cometa errores lógicos en sus razonamientos impropios de un sujeto caritativamente racional.

Conviene examinar con más detalle este argumento. Para ello, comenzaremos describiendo qué clase de fallo se produce en el auto-conocimiento de un sujeto a consecuencia de un caso de traslado de acuerdo con Boghossian. Este fallo consiste en una violación del Principio de Transparencia del Contenido, el cual establece como un rasgo de nuestros contenidos mentales que podamos saber si cualquiera de ellos es idéntico o no a cualquier otro.¹⁵² En segundo lugar, Boghossian describe el contenido de las tesis externistas diciendo que, de acuerdo con ellas, la fijación de un contenido es funcional a “la propiedad del entorno que constituye la causa típica de las percepciones que causan y sustentan” dicho contenido (1992, p. 19). Si esta propiedad del entorno cambiara sin que el sujeto que piensa el contenido en cuestión lo supiera – tal y como ocurre en los casos de traslado –, entonces dicho sujeto sería incapaz de advertir la diferencia que existe entre los contenidos que expresaba en el pasado y los que expresa en el presente mediante preferencias homófonas. Por estas razones, dado el externismo, los casos de traslado conllevan una violación del Principio de Transparencia del Contenido.

El segundo paso del argumento de Boghossian es mostrar que la incapacidad del protagonista de los casos de traslado para identificar sus contenidos da lugar a que este hablante cometa ciertos errores lógicos en su razonamiento, en concreto, falacias de equivocación. A este respecto conviene acudir al ejemplo que el autor propone en (1994, pp. 39-45).¹⁵³ Supongamos que durante su estancia en la Tierra, Óscar₁ viajó a Nueva

¹⁵² Boghossian afirma que el Principio de Transparencia del Contenido se deriva del Principio de Transparencia del Significado asumido por Frege y por otros filósofos del lenguaje. De este modo, Frege se apoya en el Principio mencionado para postular la existencia del sentido como una dimensión semántica distinta de la referencia. Según el Principio de Transparencia del Significado, todo hablante que entienda dos enunciados es capaz de saber si tienen o no el mismo sentido. Dado esto, pensemos en un sujeto que entiende las oraciones “Cristiano Ronaldo es un egoísta” y “CR9 es un egoísta” – oraciones que predicán la misma propiedad del mismo individuo – y sólo acepta la verdad de la primera y niega la verdad de la segunda. Suponiendo que un individuo racional no puede mantener actitudes contrarias hacia dos oraciones que él entiende como sinónimas, entonces (i) las oraciones “Cristiano Ronaldo es un egoísta” y “CR9 es un egoísta” no podrán ser consideradas sinónimas, (ii) los nombres propios “Cristiano Ronaldo” y “CR9” habrán de tener distinto sentido o significado y (iii) como estos nombres son correferenciales, entonces su sentido o significado debe ser distinto de su referente. Cf. Dummett (1993), p. 95.

¹⁵³ Hemos elegido este ejemplo y no el que Boghossian expone en (1992), p. 22, porque este último no sirve para impugnar la comprensión del externismo que nos interesa, a saber, la tesis que afirma que para individuar los contenidos de los pensamientos que los individuos expresan haciendo uso de ciertos términos generales es necesario contar, en ocasiones, con factores externos a sus mentes. El ejemplo que Boghossian formula en (1992) tiene la siguiente forma: (1) quiero abrazar a cualquiera que nadó una vez en el Lago Taupo; (2) Pavarotti nadó una vez en Lago Taupo; (3) por tanto, quiero abrazar a Pavarotti. En

Zelanda y vio a Luciano Pavarotti flotando en las aguas del Lago Taupo. Imaginemos que, años después, Óscar₁ es víctima de un traslado inadvertido a la Tierra Gemela. Transcurrido el tiempo necesario para que este sujeto pase a expresar un concepto distinto mediante sus usos del término “agua”, supongamos que Óscar₁ lleva a cabo el siguiente razonamiento:

(R1) Todo lo que flota en el agua está húmedo.

(R2) Pavarotti una vez flotó en el agua.

(R3) Pavarotti una vez estuvo húmedo.

De acuerdo con Boghossian, la individuación externista de los contenidos de Óscar₁ implica, por un lado, que el pensamiento que este agente expresa al proferir (R1) es un pensamiento que versa sobre el agua₂ y, por otro lado, que el pensamiento expresado mediante (R2), al ser una rememoración de un pensamiento adquirido en la Tierra, posee un contenido que versa sobre el agua. Ahora bien, esto supone que Óscar₁ está cometiendo una *falacia de equivocación* al razonar de esa manera, un tipo de error lógico que tiene lugar cuando un término que aparece en varias premisas de un argumento posee diferentes sentidos en cada una de ellas o, lo que es igual, cuando el término en cuestión expresa diferentes conceptos en las distintas premisas del argumento.¹⁵⁴

El tercer paso del argumento incompatibilista que propone este autor consiste en mostrar cómo la clase de errores lógicos que comete Óscar₁ (propiciados, no lo olvidemos, por la individuación externista de sus contenidos mentales y la consiguiente violación del Principio de Transparencia del Contenido) son impropios de un sujeto racional. La argumentación de Boghossian discurre de la siguiente manera. Nuestras atribuciones mentales ordinarias deben regirse por el Principio de Caridad, un principio que goza de aceptación general y que exige considerar a los sujetos de atribución como sujetos racionales. Esto significa que si atribuimos a un hablante la creencia de que P y sus actuaciones son contrarias a las de una persona que mantiene esta creencia, el último recurso para volver inteligible su comportamiento es cuestionar su racionalidad y atribuirle la creencia de que P y la creencia de que no P. Por lo general, antes de atribuir

este ejemplo sólo se da una falacia de equivocación si consideramos que la individuación externista se aplica a los conceptos expresados por términos singulares como “Lago Taupo” o “Pavarotti”.

¹⁵⁴ Podemos poner un ejemplo donde se cometa este error de manera flagrante: (S1) los doctores saben cómo curar enfermedades; (S2) Silvia espera que su tesis le permita ser doctora; por tanto, (S3) Silvia espera que su tesis le permita saber cómo curar enfermedades. Es obvio que el término “doctor” tiene diferentes significados en (i) y (ii) y, por tanto, este razonamiento no es válido. Lo mismo ocurre, según Boghossian, en las premisas (R1) y (R2) del razonamiento de Óscar.

creencias contradictorias resulta mucho más razonable suponer que nos hemos equivocado al adscribirle una de esas creencias, o que este sujeto ha cambiado de opinión, o que ha decidido actuar contrariamente a sus creencias obligado por alguien, etc.

La concreción que propone Boghossian del Principio de Caridad supone establecer una ecuación entre la racionalidad del sujeto de atribución y la capacidad de éste de adecuarse *a priori* a las leyes de la lógica en sus razonamientos.¹⁵⁵ Así, un hablante será considerado racional si es capaz de conocer mediante reflexión *a priori* que sus razonamientos son válidos y detectar, igualmente *a priori*, la existencia de falacias en sus propios razonamientos. Ahora bien, Boghossian argumenta que la violación del Principio de Transparencia del Contenido que se desprende de las tesis externistas da lugar a que Óscar₁ cometa errores lógicos al razonar que no puede evitar reflexionando *a priori*. Esto es, la violación de este Principio implica que Óscar₁ no puede saber *a priori* que el pensamiento expresado mediante la preferencia de (R1) es un pensamiento distinto del que expresa al emitir (R2) y, por tanto, Óscar₁ será incapaz de averiguar, asimismo *a priori*, que está cometiendo una falacia de equivocación en su razonamiento. Óscar₁ sólo se percataría de su error si le proporcionáramos información de carácter empírico sobre la distinta composición química de los líquidos relevantes. De este modo, concluye Boghossian, el externismo atenta contra el Principio de Caridad por la vía de renunciar al Principio de Transparencia del Contenido, pero como el Principio de Caridad es altamente plausible, es la individuación externista la tesis que debe ser rechazada.¹⁵⁶

Una vez expuesto el argumento de Boghossian, querríamos hacer una observación al respecto, a saber, que la explicación del cambio que se produce en el repertorio conceptual de la víctima de traslado es, en este caso, una explicación por

¹⁵⁵ Cf. Boghossian (1992), p. 17 y (1994), p. 41.

¹⁵⁶ Tanto en (1992) como en (1994) Boghossian desarrolla un argumento incompatibilista paralelo a partir de la distinción entre atribuciones *de re* y atribuciones *de dicto*. Según este autor, el externismo mantiene que las atribuciones de pensamientos interpretadas *de dicto* atribuyen pensamientos individuados en función de factores externos al sujeto. Si, por definición, la atribución de pensamientos interpretados *de re* supone atribuir pensamientos individuados en función de los objetos acerca de los que versan, el externismo conlleva que la distinción entre atribuciones *de dicto* y *de re* colapsa. El problema es que usualmente se acepta que las atribuciones interpretadas *de re* son inapropiadas para dar cuenta del comportamiento de los agentes y asegurar su racionalidad, tal y como convinimos en la introducción (cf. Owens 1994, p. 524). Al igual que en los puzzles fregeanos, un agente puede asentir ante la preferencia de “Joan Crawford fue una famosa actriz de cine mudo” y negar la verdad de “Lucille Fray Le Sueur fue una famosa actriz de cine mudo”, aun cuando estas oraciones prediquen la misma propiedad del mismo individuo. Si efectuéramos una lectura *de re* de los pensamientos atribuidos a este hablante, tendríamos que decir de él que cree y no cree que el mismo individuo posee la misma propiedad, algo que contradice nuestra suposición de que el sujeto que recibe las atribuciones es un individuo racional.

suplementación; este autor supone que el cambio de entorno se salda con la adquisición de un nuevo concepto, el concepto de agua₂, que figura en el pensamiento expresado en la premisa (R1), sin que ello implique la pérdida del concepto terrestre, que figura en el pensamiento expresado por la segunda premisa (R2). El uso de la explicación por suplementación conceptual supone una novedad en relación a su argumento de (1989), visto que la tercera premisa de éste establecía que Óscar₁ *no podía* recordar en la Tierra Gemela el pensamiento sobre el agua que mantuvo en la Tierra al carecer del concepto relevante.¹⁵⁷

En lo que sigue tomaremos en consideración las distintas objeciones compatibilistas al argumento de Boghossian basado en el razonamiento. En primer lugar, examinaremos la propuesta de Falvey en (2003). En segundo lugar, indicaremos de qué manera la memoria preservativa ofrece a Burge herramientas para enfrentarse al argumento incompatibilista basado en el razonamiento. A continuación, atenderemos a las interesantes sugerencias de Grimaltós en (2003a), basadas en la distinción entre usos deferenciales y referenciales de los términos que figuran en los informes de auto-atribución de pensamientos. Por último, analizaremos la posición de Moya de (2003), quien propone una comprensión del externismo más sofisticada que aquella que subyace al argumento de Boghossian y que nos permitirá, según nos dice su autor, evitar la comisión de las falacias de equivocidad que denuncia el incompatibilismo.

Por lo que respecta a Falvey, el eje de su argumentación se sitúa en la formación de conceptos de tipo amalgama o disyuntivo en las situaciones de traslado. Esto significa que, tras haber permanecido en la Tierra Gemela el tiempo necesario para que las condiciones del nuevo entorno jueguen un papel en la individuación de sus conceptos, Óscar₁ expresará mediante el uso del término “agua” un concepto con una extensión más amplia que la que tenía el concepto que expresaba en la Tierra. El nuevo concepto se aplica tanto al H₂O como al XYZ, mientras que el antiguo concepto sólo se aplicaba a la primera de estas sustancias. Falvey afirma que esta tesis se ve reforzada por la situación de confusión en la que se encuentra Óscar₁ tras el traslado. De este modo, dado que Óscar₁ es incapaz de distinguir entre los líquidos relevantes, sería

¹⁵⁷ Las réplicas compatibilistas basadas en la suplementación conceptual (Burge y Gibbons) son más apropiadas para responder al argumento de Boghossian sobre la memoria que las respuestas basadas en la sustitución conceptual (Ludlow y Tye). La razón es que esta última clase de respuestas legitimaba fallos sistemáticos en la memoria. Sin embargo, los autores que se adhieren a la “explicación externista de la memoria” pueden esquivar sin problemas el argumento incompatibilista basado en el razonamiento, dado que, según esta explicación, el término “agua” sería unívoco en todas las premisas de dicho argumento (adquiriría una interpretación gemelo-terrestre) y ello evitaría la comisión de una falacia de equivocidad.

imprudente interpretar de manera directa sus pensamientos en tanto que pensamientos acerca del agua₂ (Falvey 2003, p. 228).

Asimismo, este autor ofrece un argumento adicional que demuestra la superioridad de su tesis frente a aquellos autores que o bien afirman que Óscar₁ pierde el antiguo concepto terrestre (Ludlow y Tye) o bien sostienen que este sujeto retiene ese concepto, aunque los usos del término “agua” después del traslado expresen dos conceptos distintos (Gibbons). Las consideraciones de Falvey se apoyan tanto en el externismo natural basado en las ideas de Putnam como en las tesis del externismo social de Burge. Así, la primera variedad del externismo establece que a veces es necesario contar con las propiedades subyacentes de los géneros naturales del entorno externo involucrados en la historia causal de un agente para individuar sus conceptos y contenidos. Ahora bien, puesto que tanto el agua como el agua₂ han formado parte de la historia causal de Óscar₁, ambas sustancias habrán participado en la misma medida en la fijación del concepto que Óscar₁ expresa mediante el uso del término “agua”.¹⁵⁸ En lo que se refiere a la segunda comprensión del externismo, es más bien el uso que una comunidad lingüística establece para un término lo que determina, en parte, qué expresa un miembro de esa comunidad mediante el uso de dicho término. En el caso de Óscar₁ encontramos que este hablante ha delegado su uso del término “agua” en dos comunidades lingüísticas diferentes (en la comunidad lingüística terrestre y en la gemelo-terrestre). Por ello, sostiene Falvey, ambas comunidades tienen, con respecto a Óscar₁, el mismo peso en la asignación de un valor semántico al término “agua”, razón por la cual la extensión del concepto que Óscar₁ expresa mediante este término es de carácter disyuntivo: H₂O o XYZ.¹⁵⁹ Dicho todo esto, podemos apreciar cómo la

¹⁵⁸ Esta observación de Falvey se sitúa en el contexto de su discusión de las tesis de Gibbons (Falvey 2003, pp. 230-232). Lo que el primero alega es que en un caso donde un individuo ha entrado en contacto causal con dos entidades distintas (o dos géneros de entidades diferentes) sin advertir la diferencia entre ellas, sería arbitrario elegir una de ellas como la entidad que determina las condiciones de verdad de las creencias del individuo. Acudiendo al ejemplo de Gibbons que discute Falvey, si Frank confunde a dos gemelos, Vinnie₁ y Vinnie₂, y piensa que se trata de la misma persona, ¿cuál de estos individuos figuraría en las condiciones de verdad de la preferencia de Frank “Vinnie siempre va vestido de un modo muy pulcro”? Sería inapropiado escoger sólo a uno de ellos si ambos sujetos han estado involucrados en la misma medida en la historia causal de Frank. De la misma forma, tanto el agua como el agua₂ figuran en la historia de encuentros causales de Óscar₁ y, “dada la intención de [Óscar₁] de expresar un concepto unívoco mediante la palabra ‘agua’, se abre la puerta [...] a una interpretación de todas sus preferencias en tanto que preferencias que expresan el concepto disyuntivo” (Falvey 2003, p. 232).

¹⁵⁹ Este segundo argumento a favor de la interpretación disyuntiva del concepto que Óscar₁ expresa mediante sus usos del término “agua” puede entresacarse de las críticas de Falvey a la solución aportada por Tye en (1998). Como vimos, este autor afirmaba que el concepto relevante de Óscar₁ debía ser interpretado como el concepto de agua₂ en tanto que este individuo delega, tras sufrir el traslado, en los expertos de la Tierra Gemela. Sin embargo, Falvey se basa en una observación de Heal en (1998) para señalar que Óscar₁ no sólo delega en ese conjunto de expertos, ya que es muy probable que para

formación de conceptos amalgama en la víctima de traslado que propone Falvey permite soslayar las alegaciones incompatibilistas. Desde el momento en que el término “agua” proferido por Óscar₁ expresa un concepto disyuntivo al realizar el razonamiento relevante en la Tierra Gemela, podemos negar que el término “agua” sea equívoco y, por tanto, que tenga un valor semántico distinto en las premisas (R1) y (R2).

En lo que concierne a Burge, en (1998) este autor ofrece una serie de sugerencias para contrarrestar el razonamiento de Boghossian. Lo primero que conviene señalar al respecto es que su posición resulta un tanto especial. Por un lado, su idea de los conceptos amalgama podría servir, por sí misma, para elaborar una réplica al argumento incompatibilista que estamos examinando. En este sentido, Burge defiende que el tipo de transformación conceptual a la que da lugar el traslado (formación de un concepto disjunto o de un concepto amalgama) depende de “cómo el individuo está comprometido con los estándares de las comunidades en los dos entornos” (1998, p. 352). En concreto, este autor sostiene que si la víctima del traslado no tiene un fuerte compromiso con los estándares de una comunidad lingüística determinada y carece de los recursos necesarios para distinguir un género natural de otro, lo más probable es que el traslado propicie la formación de un concepto con una extensión ampliada. Tal es, como puede verse, la situación de Óscar₁, por lo cual es perfectamente concebible que el caso de traslado que contempla Boghossian se salde con la formación de un concepto amalgama, lo que permitiría evitar la equivocidad del término “agua” en el argumento de Boghossian y el consiguiente error lógico. Pero, por otra parte, Burge ofrece un argumento adicional. Según este autor, el problema que nos encontramos en los razonamientos relevantes de Óscar₁ no es un problema de carácter lógico (y resoluble *a priori* por un sujeto racional), sino que una de las presuposiciones de carácter empírico de ese razonamiento es falsa. De ser así, la individuación externista de los conceptos no atendería en absoluto contra la capacidad de los sujetos racionales de detectar *a priori* errores lógicos en sus razonamientos.

El desarrollo del argumento de Burge comienza por hacer hincapié en la consideración del sujeto de traslado como un sujeto racional. Supongamos, al igual que Boghossian, que las facultades racionales de Óscar₁ funcionan a la perfección. Es de suponer, asimismo, que este hablante tiene la intención de estar usando el término

identificar el referente del término “agua” en la Tierra Gemela este individuo emplee descripciones como “el líquido en el que nadaba cuando era un niño” que apelan al conjunto de expertos de su entorno de origen.

“agua” de manera unívoca. ¿Por qué sostiene Boghossian que la defensa del externismo implica frustrar esa intención de Óscar₁? La razón es, según Burge, que la concepción del externismo que pone en juego Boghossian es inapropiada. Así, este autor parece atribuir al externismo la tesis de que los contenidos de *todos* nuestros pensamientos están fijados *en exclusiva* por el tipo de objetos que ocasionan dichos pensamientos. De acuerdo con esto, si Óscar₁ interacciona causalmente en dos ocasiones diferentes con géneros naturales distintos (tal y como ocurre en el argumento de Boghossian), el externista debería decir que Óscar₁ mantiene contenidos diferentes en cada ocasión. Ahora bien, Burge sostiene que es posible mantener un externismo robusto que, sin embargo, también dé cabida a factores de individuación diferentes de las entidades del entorno externo que causan los pensamientos, de modo tal que las intenciones de un sujeto, por ejemplo, podrían intervenir de pleno derecho en la fijación de sus contenidos, aunque sean factores internos. De ser así, y puesto que la consideración caritativa de Óscar₁ contempla como una “exigencia de su razonamiento” su intención de usar el término “agua” unívocamente, el externista no tiene por qué rechazar la intervención de esas intenciones en la individuación de los contenidos de sus pensamientos.

El segundo paso del argumento de Burge consiste en apelar a la memoria preservativa para dar fundamento a las intenciones de univocidad de Óscar₁. En concreto, este autor declara que “uno sólo puede evitar [los errores en el razonamiento derivados de los casos de traslado] confiando firme e intencionadamente en que la memoria preservativa mantendrá los mismos conceptos a través del curso de los propios argumentos” (Burge 1998, p. 368). La memoria preservativa es, como indicamos en la sección precedente, la facultad responsable de traer al presente contenidos adquiridos en el pasado. Esta facultad también interviene en la formación de la clase de razonamientos que Boghossian examina, a saber, razonamientos en cuyas premisas figuran pensamientos rememorados. La memoria preservativa es responsable, para Burge, de que sea uno y el mismo concepto el que figura en las premisas del razonamiento de Óscar₁. Para apreciar de qué modo opera esta facultad en los casos que estamos examinando, vamos a describir un ejemplo similar al que Burge ofrece en (1998, p. 366) para extraer el esquema de los razonamientos de las víctimas de traslado:¹⁶⁰

¹⁶⁰ Lo único que hemos modificado es el nombre del protagonista del ejemplo original de Burge y los géneros naturales involucrados (aluminio y aluminio₂ en vez de agua y agua₂).

(BR1) Óscar₁ recuerda haber bebido agua.

(BR2) Óscar₁ recuerda que ayer tuvo ante sí un vaso de agua.

Por tanto, Óscar₁ concluye que una vez bebió el mismo líquido que contenía el vaso que ayer tuvo ante sí.

Los razonamientos que son objeto de atención por parte de Boghossian incluyen como primera premisa un pensamiento adquirido antes de que se haya producido el traslado, de tal manera que el concepto expresado por el término relevante sea, presumiblemente, un concepto terrestre (concepto de agua). En la otra premisa debería figurar un pensamiento adquirido después del traslado. Según Boghossian, los parámetros externistas de individuación determinan que el concepto que expresa el término “agua” en (BR2) sea el concepto gemelo-terrestre pero, puesto que esta diferencia en los conceptos es indetectable de manera puramente *a priori* para la víctima de traslado, entonces la conclusión incompatibilista está servida. El externista debería reconocer que este hablante es incapaz de evitar *a priori* la comisión una falacia de equivocidad en su razonamiento, con el consiguiente atentado contra su racionalidad.

Ahora bien, Burge cuestiona que el externista esté obligado a admitir que el término “agua” expresa en la segunda premisa un concepto distinto del que expresa en la primera de ellas. Tal obligación existiría sólo si admitiéramos la caracterización que Boghossian ofrece del externismo, de acuerdo con la cual el objeto (agua₂) que *causa* el pensamiento que Óscar₁ mantiene en (BR2) es de un tipo distinto de aquel objeto (agua) que *causaba* el pensamiento terrestre que rememora en (BR1). En este caso, deberíamos concluir, con Boghossian, que según el externismo el término “agua” debe expresar un concepto distinto en cada premisa. Sin embargo, ya hemos visto que esta concepción del externismo resulta altamente cuestionable, por lo que Burge considera que desde una concepción más apropiada de estas tesis es posible rechazar que el término “agua” expresa el concepto gemelo-terrestre en la segunda premisa. De este modo, cabe apelar a la intervención de la memoria preservativa como la facultad que permite remitir anafóricamente al concepto relevante de la primera premisa desde la segunda. Si es el concepto terrestre de agua el que figura en ambas premisas, entonces Óscar₁ no estará cometiendo una falacia de equivocación en su razonamiento. Es cierto, admite Burge, que su argumento contiene una premisa falsa, (BR2), y, por tanto, daría lugar a una conclusión falsa. Sin embargo, un error de esa clase no concierne a la validez lógica del argumento, que es la clase de problemas que Boghossian imputaba al externismo. Lo único que ocurre es que una de las presuposiciones de carácter empírico que apoyan el

razonamiento de Óscar₁ no es verdadera, ya que este sujeto está equivocado al suponer que el líquido que ayer tuvo ante sí era agua. El externista, en consecuencia, puede mantener que el error que Óscar₁ comete al razonar es un error meramente empírico y, por tanto, un error que los sujetos racionales no tienen por qué ser capaces de descartar *a priori*.

En resumidas cuentas, la réplica de Burge a Boghossian se articula en torno a dos elementos. En primer lugar, denuncia la comprensión del externismo que pone en juego Boghossian y defiende que las intenciones de un hablante al efectuar un razonamiento también cuentan como factores de individuación de los contenidos de sus pensamientos. En segundo lugar, la memoria preservativa permite mantener la coherencia lógica de aquellos razonamientos que involucran pensamientos adquiridos en el pasado e imponer un carácter unívoco a los términos usados en el razonamiento. Por todo ello, Burge se ve legitimado para descartar la existencia de errores lógicos (susceptibles de ser excluidos *a priori* por un sujeto racional) en el razonamiento de Óscar₁. El hecho de que ese razonamiento dé lugar a una conclusión falsa se debe a un mero error en sus identificaciones empíricas.

Tras describir la argumentación compatibilista de Burge, dedicaremos las próximas páginas a examinar las propuestas de Grimaltós (2003a) y Moya (2003) para resolver el problema planteado por esta versión de los casos de traslado. En lo que se refiere al primero de estos autores, Grimaltós alega que tales casos podrían interpretarse como situaciones en las que un hablante piensa acerca de (o se refiere mentalmente a) un conjunto de entidades distinto de aquél que constituye la referencia semántica de sus palabras. Grimaltós intenta dar sentido a esta clase de fenómenos a partir de la distinción entre usos referenciales y usos deferenciales de nuestros términos. En particular, esta dicotomía sirve para explicar que, en ocasiones, los contenidos mentales de la víctima de traslado pueden recibir una individuación distinta (a partir de factores naturales terrestres) de la que recibe el contenido semántico de las oraciones que él usa para expresar sus pensamientos (a partir de factores gemelo-terrestres). Según Grimaltós, puesto que es posible satisfacer la pretensión de un sujeto de remitir referencialmente a un entorno distinto del actual (aun cuando la referencia semántica de sus términos, deferencialmente establecida, apunte al entorno actual), entonces también será posible conservar su acceso privilegiado. De este modo, siempre que la víctima de traslado sepa qué referencia tienen sus pensamientos, este hablante sabrá qué es lo que piensa, sin

perjuicio de que las expresiones lingüísticas que usa para conocer dichos pensamientos no sean las apropiadas.

A fin de comprender en profundidad la estrategia de Grimaltós, expondremos a continuación su aspecto central, a saber, la distinción entre usos lingüísticos deferenciales y referenciales. Por lo general, se asume que podemos identificar el contenido proposicional de nuestros pensamientos con el contenido proposicional que expresan aquellas oraciones empleadas para informar a los demás de tales pensamientos.¹⁶¹ Para Grimaltós esto ocurre cuando los términos son usados de *modo deferencial*, ya que en esta clase de usos la identidad de nuestro contenido mental vendría determinada por el significado convencional de los términos que figuran en aquella preferencia que empleamos para expresar nuestro pensamiento. Por otro lado, si un agente comunicara uno de sus pensamientos dando un *uso referencial* a su informe, entonces el contenido de ese pensamiento no tendría por qué ser idéntico al contenido semánticamente expresado por la preferencia relevante, esto es, al contenido expresado por la preferencia usada deferencialmente.

Para aportar claridad a estas ideas resulta conveniente recurrir a la distinción entre *referencia semántica* y *referencia del hablante*, propuesta por Kripke en (1977),¹⁶² y que constituye, según nos dice Grimaltós, la inspiración de su propuesta. Por una parte, la *referencia semántica* de una expresión consiste en el objeto o estado de cosas designado en función del significado que está convencionalmente asociado a esa expresión. Sin embargo, en un momento dado un hablante que hace uso de esa

¹⁶¹ De este supuesto se desprende el siguiente principio rector de nuestras atribuciones psicológicas: siendo “vb que S” una oración en la que en el lugar de “vb” figura un verbo de actitud mental y en el lugar de “S”, una oración asertiva:

(P-At): Si un individuo profiere sinceramente la oración “vb que S” o asiente sinceramente ante su preferencia por parte de un tercero, entonces podemos atribuir a dicho individuo un pensamiento correspondiente cuyo contenido será la proposición expresada por la oración “S”.

En el caso de la actitud mental que es la creencia, puede enunciarse una concreción de este principio:

(P-At-C): Si un individuo profiere sinceramente la oración “S” o asiente ante la preferencia de “S” por parte de un tercero, entonces podemos atribuirle una creencia cuyo contenido será la proposición expresada por “S”.

¹⁶² Ya en (1957, p. 484) Grice señaló la diferencia entre lo que un hablante dice y lo que un hablante quiere decir o implica al decirlo. Por otro lado, el análisis russelliano de las descripciones definidas es, según Grice, apropiado para dar cuenta de las condiciones de verdad de oraciones que involucran descripciones definidas, tanto en sus usos referenciales como en sus usos atributivos. De este modo, pensemos en un hablante que emite la preferencia “el asesino de Smith está loco” señalando a un individuo (a la vista de su audiencia) que no es el asesino de Smith pero que muestra síntomas evidentes de locura. Esta oración, de acuerdo con Russell, sólo sería verdadera en caso de que quienquiera que sea el único individuo que verdaderamente asesinó a Smith esté loco. Si el verdadero asesino de Smith resultara estar cuerdo, el hablante *habría dicho* una falsedad, aunque lo que él quiso decir o comunicar a su audiencia (que *ese* individuo que todos creían el asesino de Smith está loco) fue algo distinto de la falsedad dicha.

expresión puede querer decir (“means”) algo distinto de lo que ese término significa (convencionalmente). En este último caso, la referencia del término así determinada sería la *referencia del hablante*.¹⁶³

Esta distinción permite entender mejor la propuesta de Grimaltós y comprobar en qué medida supone una ruptura con respecto a la identificación usual entre, por un lado, los contenidos proposicionales que figuran en nuestros pensamientos o (C-M) y, por otro lado, los contenidos proposicionales expresados por nuestras preferencias o (C-P). Asumiendo que el significado convencional de los términos que figuran en una preferencia fija el contenido (C-P) y teniendo en cuenta que ese mismo significado determina qué conceptos figuran en el contenido (C-M) cuando nuestras expresiones son usadas deferencialmente, entonces podemos establecer *la identidad de los contenidos (C-P) y (C-M) en los usos deferenciales*.¹⁶⁴ No obstante, en los usos referenciales puede ocurrir que un sujeto emplee una expresión desviándose de las convenciones semánticas asociadas a ésta. En este caso, el hablante trataría de comunicar un pensamiento acerca de una entidad o estado de cosas empleando una expresión “inapropiada”, es decir, una expresión en la cual la referencia semántica de sus términos componentes no coincide con la referencia de los conceptos que figuran en su pensamiento. Como consecuencia de ello, *en un uso referencial existiría una ruptura entre el contenido (C-P), fijado por el valor semántico convencional de los términos*

¹⁶³ Esta aportación de Kripke está basada en la distinción de Grice y permite ofrecer una lectura no semántica de Donnellan entre usos atributivos y referenciales de las descripciones definidas. La interpretación semántica de esa distinción, de la que Donnellan parece desmarcarse en (1966, p. 99), vendría a sostener que cada descripción definida posee dos significados, que se corresponden con cada uno de los usos, y establecería distintas condiciones de verdad para cada uno de ellos. Kripke defiende la siguiente interpretación de la distinción entre usos atributivos y referenciales. Mientras que en un uso atributivo de la descripción definida “el F” no hay diferencia entre el significado literal de dicha descripción (referencia semántica) y lo que el hablante quiere decir mediante su uso (referencia del hablante), en un uso referencial lo que un hablante trata de comunicar puede ser distinto de lo que la descripción expresa según su significado léxico. Esto se aprecia en el caso de usos referenciales de descripciones definidas que seleccionan a un individuo que no satisface la propiedad expresada por la descripción. Si le digo a mi interlocutor “el vecino del quinto me ha denunciado” mientras señalo al vecino del cuarto piso, se pone de manifiesto una divergencia entre lo que digo y lo que pretendo comunicar a mi amigo. Kripke establece la compatibilidad entre la teoría russelliana de las descripciones definidas y la propuesta de Donnellan del modo siguiente. Mientras se reserve el uso referencial de las descripciones definidas como una cuestión propia de la pragmática (relacionada con las intenciones comunicativas del hablante) y consideremos que los distintos usos de una descripción no suponen una diferencia en su semántica, podemos considerar que el análisis de Russell es correcto en lo que concierne a la semántica de (todos los usos de) las descripciones definidas.

¹⁶⁴ Esta clase de uso es el que subyace a la lectura anti-individualista del experimento mental de Burge de (1979). De este modo, consideramos que existe una diferencia entre los contenidos que Bert y Bert-contrafáctico expresan al decir “tengo artritis en el muslo” porque (i) la referencia semántica de “artritis” difiere en sus respectivas comunidades y (ii) asumimos que ambos están haciendo un uso deferencial del término.

contenidos en la preferencia, y el contenido (C-M), determinado por los conceptos que contiene el pensamiento del hablante.

Veamos de qué modo la distinción entre el uso referencial y el uso deferencial da respuesta a la lectura incompatibilista de los casos de traslado. La propuesta de Grimaltós tiene como objetivo respetar nuestra intuición de que “es la intención del hablante la que fija la referencia de sus términos” (2003a, p. 36). De este modo, aunque el traslado inadvertido conlleva un cambio en la referencia semántica del término “agua” (dado que, según Grimaltós, esa referencia pasa a ser el XYZ), esto no significa que Óscar₁ no pueda pensar ocasionalmente en el H₂O, aunque en la preferencia que usa para expresar sus pensamientos ninguno de los términos componentes designe semánticamente ese líquido.

El diagnóstico del autor sobre la clase de razonamientos que contempla Boghossian en su trabajo guarda cierto parecido con la propuesta de Burge de (1998). Su idea es apelar, de nuevo, a las intenciones del hablante, quien pretende estar expresando el mismo concepto en todas las ocurrencias del término “agua” en cada una de las premisas del razonamiento. Así, el recurso de separar contenidos lingüísticos y contenidos mentales sirve para evitar la comisión de una falacia de equivocidad en el argumento de Óscar₁ o, en el peor de los casos, restringir esta equivocidad al ámbito lingüístico. De acuerdo con esto, aunque la víctima de traslado expresa verbalmente su razonamiento usando una premisa que versa sobre el agua y otra que versa sobre otra clase de líquido, esta diferencia no se trasmite al ámbito de sus pensamientos. La consideración caritativa de la racionalidad de Óscar₁ exige univocidad en sus pensamientos y no hay ninguna razón por la que el externista deba frustrar esa pretensión.

Visto esto, podemos apreciar que la posición de Grimaltós pretende conjugar de manera coherente tres factores de individuación de contenidos: la comunidad lingüística de deferencia, el entorno natural que constituye el origen causal de nuestros pensamientos y su origen inferencial.¹⁶⁵ La idea de independizar contenidos mentales y lingüísticos sirve para equilibrar las tesis del externismo natural, por un lado, y las tesis del externismo social por otro. Grimaltós desplaza la tensión entre el externismo y el acceso privilegiado hacia un conflicto entre, por una parte, la adscripción de

¹⁶⁵ No obstante, estamos de acuerdo con Liu (2002) en que las tesis del externismo natural y las tesis del externismo social no son tan compatibles entre sí como a menudo se presume. En la sección 2.3 del quinto capítulo de este trabajo tendremos ocasión de argumentar este extremo detalladamente.

pensamientos a partir de las preferencias del hablante, que tenderá hacia la identificación entre contenidos pensados y contenidos expresados semánticamente, y, por otra parte, la intención del hablante de pensar en una clase de sustancia que no es la designada semánticamente por sus preferencias.

Aunque esta estrategia contiene aspectos muy sugerentes, estimamos que no es del todo apropiada para discutir los casos de traslado que Boghossian contempla. Tales casos se basan, como vimos, en el experimento mental en torno al término “agua” destinado a mostrar la viabilidad de una individuación externista *estrictamente en función de factores naturales*, donde no interviene en absoluto la división social del trabajo lingüístico. Recordemos, en efecto, que en 1750 no existen “expertos” con respecto a la extensión del término “agua” en ninguno de los dos entornos. Por este motivo, las condiciones de aplicación de esa expresión estarán determinadas, presumiblemente, por las propiedades macroscópicas de los líquidos, que, por hipótesis, son idénticas, ya que en ambos entornos se denominará “agua” al líquido inodoro, incoloro e insípido que corre por ríos y lagos y sacia la sed de los seres vivos. A pesar de ello, los partidarios del externismo natural cuestionan que las convenciones semánticas asociadas con el término “agua” sean idénticas en ambos entornos, algo que, de ocurrir, abriría la posibilidad de que ese término tuviera una extensión disyuntiva (H_2O o XYZ). El argumento que suele emplearse para evitar una conclusión semejante es que las normas convencionales de uso asociadas con los términos de género natural incluyen términos indécicos, que seleccionan una extensión distinta en cada entorno. Ahora bien, consideramos que el componente indécico de las convenciones semánticas asociadas con el término no tiene la fuerza suficiente como para establecer esta diferencia en las convenciones comunitarias de los dos entornos. En lo que sigue, argumentaremos a favor de esta tesis y nos apoyaremos en las observaciones de Burge (1982) y, especialmente, de Liu (2002) para reforzarla.¹⁶⁶

¹⁶⁶ Con respecto a la tesis de que existe una diferencia semántica porque la extensión es distinta en un sentido relevante, conviene indicar que las últimas secciones del quinto capítulo de nuestro trabajo están dedicadas a cuestionar la teoría de los géneros naturales de Putnam, que constituye el presupuesto fundamental para aceptar la relevancia de la distinción en las extensiones del término “agua” en un entorno y otro. En cualquier caso, independientemente de que exista una diferencia en el significado del término “agua”, el punto que defendemos es que desde la perspectiva de Putnam no es una diferencia que concierna a las convenciones asociadas comunitariamente con el término. Precisamente, la tesis que mantiene este autor es, como alega Liu, que para individuar el significado de los términos de género natural debemos contar con factores que rebasan lo que saben los hablantes, considerados de una manera individual o comunitaria.

Para justificar esta idea, tomemos en consideración una observación de Burge en relación a la presencia de indécicos en la fijación ostensiva de la referencia. Según reconoce este autor (Burge 1982, p. 89), si los habitantes de la Tierra y de la Tierra Gemela hubieran entrado en contacto, es de prever que el ámbito delimitado inicialmente por el indécico “aquí” que figura en la definición ostensiva del término “agua” se habría ampliado, hasta el punto de incluir tanto *nuestros* ríos y lagos como los *suyos*.¹⁶⁷ La razón es que, atendiendo a los términos en que está formulado el experimento de Putnam, por lo que respecta a los hablantes de la Tierra y de la Tierra Gemela los líquidos en cuestión son todo lo similares que cabría imaginar, puesto que sólo difieren en una propiedad que *nadie* conoce. Esto supone que el poder de restricción a un entorno que tiene el componente indécico de dos definiciones ostensivas verbalmente idénticas (como las que asocian las comunidades lingüísticas de la Tierra y de la Tierra Gemela con el término “agua”) ha de relativizarse, en atención a la movilidad histórica y geográfica de una comunidad de hablantes.¹⁶⁸

En primer lugar, Liu mantiene en (2002) que el objetivo de Putnam al plantear el escenario del experimento mental en torno al uso del término “agua” en 1750 no es únicamente reforzar el eslogan de que “los significados no están en la cabeza” de los hablantes individuales, sino ampliar su alcance arguyendo que tampoco se encuentran en el estado mental colectivo de una comunidad lingüística. Por este motivo, al postular que el agua y el líquido relevante de la Tierra Gemela comparten todas las propiedades que los hablantes conocen (es decir, todas las propiedades superficiales) se asegura de que en ambos entornos se asocian con el término “agua” los mismos criterios de aplicación, o, con otras palabras, que ambos entornos se encuentran en el mismo estado

¹⁶⁷ Una observación similar es hecha por Searle (1983), p. 209. Esto nos indica que autores tanto de filiación internista (Searle) como externista (Burge) han cuestionado que el carácter indécico de las definiciones ostensivas de Putnam permita por sí sólo restringir la extensión del término “agua” al H₂O.

¹⁶⁸ Encontramos la misma idea en Marconi (1997), quien señala, además, que hablar de “fundamentaciones múltiples” más bien que de un “bautismo inicial”, como proponen Devitt/Sterelny en (1999), no supone ninguna mejora en este sentido. Según estos autores, la determinación de la referencia de un término de género natural (y también de un nombre propio) no se lleva a cabo mediante un bautismo inicial, sino que esta expresión queda vinculada a su referente mediante (múltiples) usos del término que, típicamente, involucran una aplicación directa del mismo en confrontación perceptiva con las instancias del género. Sin embargo, de acuerdo con Marconi, “si uno se adhiere a la idea de las fundamentaciones múltiples, la visión objetivista [de acuerdo con la cual el término terrestre “agua” sólo se refiere al H₂O, por ser ésta la sustancia involucrada en las prácticas causales-referenciales en el uso de ese término por parte de los habitantes de la Tierra] resulta difícil de mantener en este caso. Supóngase en efecto que un grupo de gemeloterrícolas visitara la Tierra antes de 1750. Entonces el uso de “agua” (en gemeloterrícola) para designar tanto XYZ como H₂O hubiera pasado a ser incluso *fundamentado*” (Marconi 1997, p. 124).

mental colectivo cuando conocen el significado del término “agua”.¹⁶⁹ Si la conclusión que Putnam extrae del experimento consiste en que “las convenciones sociales/lingüísticas no pueden determinar completamente el significado” (Liu 2002, pp. 286 y ss.), ello se debe a que, *a pesar de ser idénticas las convenciones que ambas comunidades asocian con el término “agua”*, el significado del término difiere en un entorno y en otro. Por este motivo, parece que existen razones para pensar que no hay nada en esas convenciones semánticas asociadas en ambas comunidades con el término “agua” que pueda impedir que este término designe de manera natural tanto al H₂O como al XYZ en el momento en el que se desarrolla el experimento de Putnam, puesto que para cualquier habitante de la Tierra y de la Tierra Gemela no se trata sólo de sustancias que guardan un parecido relevante, sino que son líquidos exactamente idénticos (¡como dos gotas de “agua”!).

Lo que nos interesa resaltar es que esta conclusión atenta contra uno de los supuestos del argumento desarrollado por Grimaltós, a saber, que los usos deferenciales del término “agua” en un entorno y en otro determinan contenidos distintos. Por el contrario, creemos que, en realidad, cuando Óscar₁ y Óscar₂ profieren la oración “el agua es húmeda” la proposición o contenido semántico expresado por sus emisiones lingüísticas podría ser exactamente el mismo en ambos entornos si apelamos a la descripción de “usos deferenciales” propuesta por Grimaltós. Es más, resulta bastante paradójico considerar que un supuesto uso “deferencial” del término “agua” marcaría algún tipo de diferencia en un entorno como el descrito por Putnam en su experimento, en el que *no* existen expertos sobre el uso del término “agua” en ninguno de los dos entornos hacia los que el resto de los hablantes puedan presentar una actitud de deferencia. Si esto es así, el contenido semántico de las oraciones de estos hablantes se fijará a partir de convenciones que harán alusión a las propiedades superficiales de los líquidos (idénticas por hipótesis) y, en consecuencia, no hay nada que impida que la referencia semántica del término “agua” pueda ser la misma en principio para ambos hablantes (y sus respectivas comunidades): H₂O o XYZ. Con otras palabras, el término “agua” podría perfectamente tener una extensión disyuntiva de acuerdo con las

¹⁶⁹ Cabría objetar que un entorno no puede ser sujeto de atribuciones de estados psicológicos en un sentido literal. Sin embargo, entendemos la afirmación de Liu de que esos “individuos” que son las comunidades de la Tierra y de la Tierra Gemela se encuentran en el mismo estado mental colectivo cuando conocen el significado del término “agua” como equivalente a decir que en ambos casos se asocian colectivamente los mismos criterios de aplicación con el mismo término (“agua”).

condiciones de aplicación de la expresión comunitariamente establecidas en ambas comunidades lingüísticas.

De aceptar esta idea, podemos ver qué ocurriría en caso de que Óscar₁ y Óscar₂ hicieran un uso deferencial de sus términos al proferir la oración “el agua es húmeda”. Según acabamos de ver, ambas comunidades (la terrestre y la gemelo-terrestre) regulan mediante convenciones verbalmente idénticas el uso del término “agua”. De este modo, si suponemos que la semántica convencional del término permite cabalmente una extensión disyuntiva para el mismo, entonces Óscar₁ y Óscar₂ estarían expresando en ese caso el mismo pensamiento al decir “el agua es húmeda” al hacer un uso deferencial de los términos involucrados en dicha oración y, de ser así, la extensión del concepto determinada de esta manera será también disyuntiva H₂O o XYZ.¹⁷⁰

Con esta posibilidad en mente, podemos examinar qué pasaría si uno de estos hablantes expresara sus pensamientos usando referencialmente los términos que figuran en sus informes. Para empezar, Grimaltós afirma que esta clase de usos se producen cuando el contenido semánticamente expresado por una preferencia como “el agua es húmeda” – contenido (C-L) – no coincide con el contenido de la creencia que el hablante pretende expresar haciendo uso de dicha preferencia – contenido (C-M) –. Según Grimaltós, estos usos responden a situaciones en las que el origen causal de un pensamiento tiene más influencia en su individuación que las convenciones semánticas de la comunidad de deferencia lingüística. De acuerdo con su propuesta para enfrentarse a los casos de traslado, si Óscar₁, tras sufrir el traslado, hace un uso referencial de la preferencia “el agua del río en el que me bañaba en mi infancia estaba fría”, el contenido que este sujeto pretende expresar mediante dicha preferencia será distinto del contenido semántico de dicha oración. De este modo, en la creencia de Óscar₁ no podrá figurar el concepto relevante de extensión disyuntiva, ya que en ese caso no habría diferencia entre (C-M) y (C-L). Debe tratarse, pues, de un concepto diferente, cuya identidad viene determinada por la naturaleza de aquellas entidades causalmente responsables de su adquisición. Ahora bien, es digno de señalarse que está completamente fuera del alcance de este agente y de sus interlocutores advertir que el concepto que figura en sus creencias en el uso referencial (concepto de H₂O) es distinto del concepto que, según suponemos, figuraría en sus creencias en la situación deferencial (concepto de H₂O o XYZ). Debido a esto, podría pensarse que la motivación

¹⁷⁰ Dicho de otra manera, los conceptos usados referencialmente tendrían una individuación más fina que los conceptos usados deferencialmente.

que guía el trabajo de Grimaltós, que es respetar el acceso privilegiado de un hablante por la vía de respetar sus intenciones al usar los términos, es una motivación vacía en un caso de traslado tipo, ya que Óscar₁ ni sabe ni puede saber por todos los medios a su alcance que pretende referirse a un líquido distinto cuando hace un uso referencial del término “agua”.

A pesar de lo dicho, creemos que las propuestas de Grimaltós están bien encaminadas y que las dificultades que presenta su aplicación a esa clase de situaciones no responden tanto a sus deficiencias internas como al hecho de que los casos de traslados tipo se plantean a partir de una comprensión del externismo que debe ser rechazada. Aunque la justificación de esta idea será presentada en rigor en el último capítulo de este trabajo, podemos presentarla efectuando el siguiente recorrido. En primer lugar, examinaremos qué clase de fenómenos lingüísticos presentan un desfase localizado y controlado entre los contenidos que un hablante pretende expresar y los contenidos que sus preferencias expresan de acuerdo con su significado estándar. En segundo lugar, alegaremos que la situación de Óscar₁ al hacer un uso referencial de sus expresiones después de sufrir el traslado no puede asimilarse a los fenómenos mencionados. En tercer lugar, examinaremos las propuestas de Moya en (2003), muy cercanas a las del propio Grimaltós y, finalmente, formularemos nuestra propuesta, basada en una combinación de las ideas de estos autores.

Atendiendo a los propósitos anunciados, cabe afirmar, en primer lugar, que la divergencia entre los contenidos pensados y los contenidos semánticamente expresados por una preferencia asertiva¹⁷¹ de un hablante es un fenómeno que se presenta de manera clara en una serie de situaciones.¹⁷² Por ejemplo, un hablante puede cometer un *lapsus* y emitir una preferencia como “en ciertos países se admiten los matrimonios polígonos” aunque, en realidad, lo que quiere decir es que en ciertos países se admiten los matrimonios polígamos. Ante esta situación nos sentimos inclinados a reconocer que

¹⁷¹ Explicitamos el carácter asertivo de las preferencias para señalar que nuestro tratamiento se ciñe a preferencias dotadas de esa fuerza ilocucionaria. Es probable que el uso de la ironía, por ejemplo, se sustente en la distinción entre lo que un individuo piensa y quiere decir y lo que sus preferencias expresan de acuerdo con su significado literal. Sin embargo, sólo nos ocuparemos de las preferencias asertivas.

¹⁷² Ciertas salvedades que suelen introducirse en el principio de desentrecorillamiento responden a estas situaciones (descuido o ignorancia). Kripke dice al respecto “Cuando nosotros suponemos que estamos tratando con un hablante normal del inglés, lo que queremos decir es que él usa la oración para decir [“mean”] lo que un hablante normal del inglés diría por medio de ella... La cualificación ‘bajo reflexión’ nos previene contra la posibilidad de que un hablante pueda, bien por una falta de cuidado o desatención con respecto al significado o bien por una confusión momentánea de carácter conceptual o lingüístico, aseverar algo que realmente no quiere decir, o asentir a una oración que contiene un error lingüístico” (Kripke 1979, p. 113).

lo que este hablante estaba pensando – el contenido (C-M) – y lo que ha dicho finalmente – el contenido (C-P) – no coinciden. Otras situaciones pueden analizarse del mismo modo. Si un hablante (dotado de una competencia media en su lengua materna) aprende una segunda lengua, es natural que al comienzo del proceso de aprendizaje cometa errores que consisten en tomar equivocadamente ciertas expresiones de la segunda lengua como sinónimas de otras expresiones “nativas”.¹⁷³ Asimismo, podemos imaginar la situación de un sujeto que entabla una conversación con un agente que habla una variante dialectal de la misma lengua. En este caso, podría ocurrir que el hablante empleara una expresión que también conoce el oyente, pero en una acepción distinta.¹⁷⁴

Estos casos presentan tres características comunes. La primera es que en todos ellos cabe afirmar que el hablante ha escogido una expresión que *no es del todo apropiada* para que el oyente interprete fielmente sus pensamientos, sin perjuicio de en algunos de estos casos el propio hablante los conozca bien. La segunda característica es que, debido a esa cierta impropiedad, consideramos que se ha producido un “conflicto” en la interpretación de los términos del hablante, conflicto que puede resolverse si el hablante rectifica el uso de los términos (*lapsus* o equivocaciones entre distintos idiomas) o si acuerda o “negocia” con el oyente una interpretación distinta en base a la ambigüedad de los términos que ha empleado (ambigüedad surgida entre variaciones dialectales). Preferimos hablar de *conflicto* en la interpretación y no de error en sentido estricto porque la impropiedad detectada no puede explicarse en ninguno de estos tres casos como una *mala* interpretación del hablante por parte del oyente. Desde luego, la interpretación no discurre sin trabas porque lo que el hablante dice (de acuerdo con la interpretación recibida) no es lo que quería decir, pero hay un sentido en el que esa interpretación que ofrece el oyente es perfectamente correcta: en el primer caso, según unos estándares lingüísticos que tanto él como su interlocutor comparten y dominan; en el segundo, de acuerdo con unos estándares que el oyente domina mejor que el hablante y, en el tercero, según unos estándares que asignan significados distintos a las mismas palabras. En cualquier caso, la rectificación o negociación mencionadas podrían tener lugar, bien de manera inmediata, al ser el oyente quien caiga en la cuenta de que, por

¹⁷³ De hecho, este fenómeno es muy frecuente en el caso de los “falsos amigos”, es decir, expresiones de la segunda lengua que por su semejanza con otras expresiones de la lengua materna sugieren al hablante una equivalencia de significado.

¹⁷⁴ Quizá podríamos extender este fenómeno a las variantes diastráticas y diafásicas de un mismo idioma. Ludlow en (1995a) plantea un caso de traslado en torno al uso del término “pragmatist”, que se aplica tanto a los seguidores de la corriente filosófica denominada “pragmatismo” como a las personas carentes de principios.

descuido, ha escogido unas palabras que no significan lo él quería decir (*lapsus*), bien de manera mediata, al consultar un diccionario o al ser un segundo hablante (competente en el uso de los términos relevantes) quien advierta el error.¹⁷⁵ En todas estas situaciones, el hablante reconocería que la expresión que ha elegido no es del todo apropiada¹⁷⁶ para que el oyente pueda interpretar bien sus intenciones, lo cual presupone que el propio hablante sabe bien cuáles eran dichas intenciones.

Para ser precisos en este punto cabe introducir una tercera característica. Estas situaciones en las que se da una divergencia de contenidos (C-M) y (C-P) están bien localizadas y su reconocimiento no implica caer en una concepción del lenguaje que haga de la comunicación un puro intento de “adivinar” lo que el hablante quiere decir con independencia de las palabras que ha usado para decirlo (interlocutor Humpty Dumpty). Una tarea de esta clase, que reserve al hablante el derecho a negociar todas las interpretaciones que recibe, haría de la comunicación un intercambio imposible. Por el contrario, en las situaciones de *lapsus*, en los errores en la traducción y en las ambigüedades dialectales nos encontramos con un hablante que, aun cuando no ha dicho lo que el oyente ha interpretado de acuerdo con lo que ha expresado semánticamente, sí reconoce la *legitimidad* (total o parcial) de esa interpretación. Incluso en el último caso, puede negociar la interpretación que recibe demostrando que el contenido que él trataba de expresar es *otra* interpretación legítima de sus palabras (de acuerdo con los estándares de la variante lingüística que él habla). Las rectificaciones y el derecho a acordar otras interpretaciones justificadas son recursos amparados por marcos de normatividad lingüística que impiden que las palabras insertas en dicho marco tengan *cualquier* significado que un hablante tenga la intención de darles.¹⁷⁷

¹⁷⁵ La rectificación podría no darse si, por ejemplo, el hablante del ejemplo hubiera dicho para sí esas palabras sin prestar demasiada atención o si su interlocutor decidiera pasar por alto el error al sobreentender que quería hablar de poligamia y no de polinomios. Lo importante es que la posibilidad de rectificación queda abierta y al alcance de los participantes en el proceso de interpretación.

¹⁷⁶ La impropiedad señalada tiene que ver o bien con el hecho de que el individuo ha escogido una expresión distinta de la que debería haber escogido para que el contenido semántico y el contenido que él intenta comunicar coincidieran en la interpretación del oyente (*lapsus* y equivocaciones en la traducción) o bien con una situación en la que el hablante ha escogido una expresión apropiada según los estándares de su comunidad pero inapropiada según los estándares de la comunidad del oyente, que establecen un uso distinto para la misma.

¹⁷⁷ Es preciso realizar una aclaración en este punto. En (1986) Davidson argumentó contra la idea de que los intercambios comunicativos exitosos entre un hablante y un oyente en una ocasión dada dependen de la existencia de un lenguaje previo compartido por ambos individuos. Este autor presentó un argumento bajo la forma de una reducción al absurdo de dicha idea, inspirándose en el fenómeno de los malapropismos, esto es, situaciones en las que el hablante emplea una expresión desviándose del uso convencionalmente asignado a ésta. Así, supongamos que el hablante y el oyente comparten un lenguaje

En segundo lugar, a pesar de que Grimaltós alude en los casos de traslado a una divergencia en los contenidos (C-M) y (C-P) de Óscar₁ y a la “impropiedad” en la elección de sus expresiones,¹⁷⁸ cabe decir que la situación de *este hablante* no es plenamente asimilable a la situación de un agente que se encuentra en alguno de los tres

previo, es decir, una teoría que recoge las mismas convenciones sistemáticamente asociadas con los mismos términos. Ahora bien, imaginemos que el hablante emplea una expresión dándole un uso distinto del convencional y, sin embargo, el oyente entiende perfectamente al hablante. Para ello, se dirá, el hablante ha necesitado una segunda teoría o teoría de paso, que incluye, además de la teoría previa, el uso desviado de la expresión involucrada en el malapropismo. La lectura que Davidson extrae de esta situación es que, en último término, la teoría previa es irrelevante para el entendimiento entre estos individuos, puesto que aquello que determina el éxito comunicativo es, al fin y al cabo, la teoría de paso. Ahora bien, “la teoría de paso no es una teoría de lo que alguien (excepto quizá un filósofo) llamaría un lenguaje natural actual” (1986, p. 102) o, dicho de otra manera, no tiene nada que ver con un conjunto de convenciones sistemáticas, sino que se construye y modifica en cada intercambio comunicativo a partir de las intenciones de los intervinientes y de su contexto extralingüístico. Una vez expuesto el argumento de Davidson, nos preguntamos en qué medida su conclusión es incompatible con nuestra indicación sobre la necesidad de contar con marcos de normatividad lingüística. En nuestra opinión ese argumento toma pie (de manera nada inocente) en una clase muy concreta de malapropismos, aun cuando pretende extraer una conclusión de carácter general. En este sentido, este autor presenta a un hablante y a un oyente que (i) son competentes en el uso convencional de un término, (ii) saben qué uso no-convencional se ha dado a ese término en esa situación y, por tanto, (iii) han llevado a cabo un intercambio comunicativo perfectamente exitoso. De esta manera, podría dar la impresión de que la constrictión convencional es prescindible porque, al fin y al cabo, lo que ha determinado el éxito de la comunicación es el uso no-convencional del término en esa ocasión concreta. Sin embargo, en lo anterior no nos estamos ocupando de malapropismos en general, sino de tres clases de ruptura localizada del contenido mental y el contenido lingüístico (*lapsus*, falsos amigos y ambigüedades dialectales), cuyos resultados reciben de manera natural un diagnóstico de carácter normativo, ya sea mediante el reconocimiento del error por parte de un hablante competente que comete un *lapsus* (“lo que dije no *se dice* así”), la corrección del mal uso que un hablante extranjero da a una expresión cuando la confunde con un falso amigo (“lo que has dicho no *se dice* así”) o mediante la negociación del uso por parte del hablante de una variante dialectal de nuestra lengua (“lo que he dicho también *se dice* así”). Por otro lado, es frecuente encontrar en los textos de Davidson ciertos indicios que revelan que la concepción de la normatividad lingüística que él critica tiene un carácter casi platónico. En este sentido, él combate la idea de que dichas convenciones son lo único que determina el éxito del intercambio comunicativo. Sin embargo, no afirmamos eso. De hecho, sería absurdo no reconocer la existencia de “pistas”, contextuales o pragmáticas, como él las denomina, que en muchas ocasiones sustentan el hilo de la comunicación. De este modo, si mi interlocutor, cuya lengua nativa es el italiano, profiere la oración “necesito un poco de burro” mientras señala sus tostadas y yo sé, además, que la traducción italiana del término “mantequilla” es “burro”, es natural que le acerque la mantequilla y no un asno. Por cortesía, sin embargo, le señalaré su error para que evite situaciones complicadas en el futuro. Aun cuando el contexto de la mesa del desayuno y su propia indicación me ayuda a saber que él quiere referirse a esa sustancia (en otro contexto quizá no lo habría tenido tan claro), no carece de relevancia el hecho de que yo conozca el significado convencional en italiano del término “burro”, que es como en esa lengua *se dice* “mantequilla”. Por último, conviene hacerse eco de la observación de Talmage (1996, pp. 338-339), quien señala que el argumento de Davidson reconoce la influencia de un cierto tipo de constrictión social en los intercambios comunicativos aunque tenga un carácter más débil que aquellas que él critica (y que tampoco son las que nosotros defendemos, como acabamos de decir). En concreto, Talmage señala que el hablante del ejemplo de Davidson, que usa de manera desviada una expresión, está tomando en cuenta cómo será interpretado (previsiblemente) por el oyente, procurándole pistas, principalmente contextuales, para que pueda tomar esa expresión en el sentido no convencional que él quiere darle. Desde luego, si el hablante se preocupa del modo en que el oyente interpretará su expresión, entonces está tomando en consideración una clase de constrictión social, en tanto que se trata de una limitación que otro hablante le impone. Además, podríamos añadir que es muy probable que muchas de las pistas no verbales empleadas por el hablante para hacerse entender tengan un sentido convencionalmente regulado.

¹⁷⁸ Cuando Óscar₁, víctima del transporte, emite la preferencia “el agua del río en el que me bañaba en mi infancia estaba realmente fría”, este hablante “ha expresado una creencia verdadera sobre el agua, que expresa (o piensa) por medio de una oración inapropiada” (Grimaltós 2003a, p. 43).

casos mencionados y la razón es que la posibilidad de rectificación y/o negociación está fuera del alcance de Óscar₁. Los interlocutores con los que entra en contacto en ambos entornos y los diccionarios que puede consultar son ciegos ante su equívoco. Ni este hablante advertiría, por sus propios medios, que existe una divergencia entre lo que quería decir y lo que finalmente dijo al proferir “el agua del río en el que me bañaba en mi infancia estaba fría” (como en un *lapsus*), ni una consulta a su diccionario o al diccionario de la Tierra Gemela le ayudaría a percibir esa divergencia y, desde luego, ninguno de sus interlocutores terrestres o gemelo-terrestres tendría noticia del supuesto equívoco. En resumen, el conflicto en la interpretación de Óscar₁ resulta extremadamente peculiar porque nadie tendría noticia de él *hasta el desarrollo de la ciencia química*.

Esto nos lleva a pensar que la posibilidad de rectificación y/o negociación requiere la existencia de una diferencia semántica efectiva (perceptible por el hablante y/o sus interpretes) entre aquellos términos que figuran, por un lado, en la oración que el hablante profiere y, por otro, en la oración que éste debería haber proferido para que el oyente pudiera capturar de manera apropiada sus pensamientos. Al decir esto no estamos manteniendo que los términos en cuestión deban ser distintos: en el caso de las ambigüedades dialectales puede no existir una diferencia gráfica en los términos en cuestión (como sí ocurría en el caso de “polígamo” y “polígono”).¹⁷⁹ Lo único que se necesitaría en este caso es una asignación semántica estándar distinta para uno de los términos, que conlleve una diferencia en la interpretación comunitariamente establecida para la oración.

Lo que sostenemos de acuerdo con lo anterior es que la dicotomía entre los contenidos pensados y los contenidos semánticamente expresados mediante nuestras oraciones tiene pleno sentido en aquellos equívocos en los que existe la posibilidad de rectificar y/o negociar apelando a una divergencia en las normas semánticas que regulan el uso de los términos, algo que no ocurre en el caso que examina Grimaltós. Pensamos, además, que un partidario del externismo como él debería tener cierto interés en tener bajo control dicha dicotomía y restringirla a esta clase de fenómenos, ya que si permitiéramos una desconexión sistemática de estos contenidos, entonces

¹⁷⁹ Pongamos como ejemplo la preferencia de una oración como “cuando llegó tu madre, llegó la zorra” por parte de un hablante de Chile que, de acuerdo con los estándares lingüísticos de su comunidad, emplea el término “zorra” como sinónimo de enfrentamiento o discusión, de acuerdo con su diccionario. Si su interlocutor fuera hablante del castellano de España habría entendido algo extremadamente ofensivo, aun cuando su hablante sólo quería decir que la presencia de la madre de ese interlocutor fue el detonante de un conflicto.

bloquearíamos la lectura anti-individualista de los experimentos mentales que suelen emplearse a favor del externismo, tales como el experimento de Burge en torno al término “artritis” (Jackman 2003, p. 35).¹⁸⁰ De hacerlo así, podríamos decir que el contenido expresado por la preferencia de Bert “tengo artritis en el muslo” es distinto del contenido mental correspondiente a dicha preferencia. Según esto, como ambos contenidos reciben evaluaciones independientes (uno puede ser verdadero y el otro falso), se argüiría que el contenido de la preferencia “tengo artritis en el muslo” es falso mientras que el pensamiento correspondiente es verdadero. Lo que se desprende de esta lectura es que el concepto involucrado en el pensamiento de Bert no es el concepto de artritis (si así fuera, su pensamiento también sería falso), pero si aceptamos eso, estamos rechazando la conclusión anti-individualista de Burge. Por esta razón, un externista debería hacer un uso limitado de la separación entre contenidos mentales y contenidos semánticamente expresados mediante preferencias. Lo que proponemos es limitar esta brecha a la comisión de equívocos lingüísticos por parte del hablante que él mismo o uno de sus interlocutores podrían, en principio, detectar y corregir. Esto nos llevará, como veremos al final de la sección, a restringir la propuesta de Grimaltós a aquellos casos de traslado donde sí existe una genuina diferencia en las prácticas lingüísticas de los distintos entornos involucrados en el traslado.

A fin de alcanzar mayor claridad en nuestra propuesta y siguiendo con el programa que hemos diseñado, examinaremos, en tercer lugar, la posición de Moya en (2003).¹⁸¹ Este autor intenta impedir el establecimiento de las conclusiones de Boghossian en (1992) y (1994) denunciando que la comprensión del externismo psicológico y del fenómeno de la deferencia lingüística que subyace a estos trabajos resulta muy cuestionable.

La concepción del externismo empleada para formular los casos de traslado sostiene que la individuación externista de los contenidos mentales toma en consideración qué “propiedad del entorno [es la que] constituye la causa típica de las percepciones que causan y sustentan” dicho contenido (Boghossian 1992, p. 19). De esta manera, si un cambio de entorno conlleva un cambio en las propiedades relevantes del entorno natural con las que un sujeto interacciona causalmente, entonces el

¹⁸⁰ El propio Jackman ofrece en (2003) otras razones para deslegitimar una desconexión sistemática de esta clase y es que, de permitirla, una afirmación como “P, pero yo no lo creo” dejaría de tener un carácter paradójico, en la medida en que el contenido semánticamente expresado por “P” es sistemáticamente distinto del contenido creído por el hablante que profiere esa oración.

¹⁸¹ Grimaltós reconoce el parentesco existente entre su propuesta y la de Moya en (2003), p. 44, n. 20.

externista debe establecer una variación correspondiente en los contenidos mentales de dicho sujeto. Aceptando que la diferencia en las propiedades naturales se sitúa fuera del alcance epistémico de dicho sujeto, la variación en sus propios contenidos le resultará inadvertida y, con ello, el argumento incompatibilista estará servido. A partir de esta situación, Moya opta por rechazar esta variante del externismo, basada en las interacciones causales típicas, y, en su lugar, nos propone una comprensión de esta tesis más sofisticada y más acorde, en su opinión, con las pretensiones teóricas de Putnam y Burge. Para mostrar el contraste entre estas dos versiones del externismo, enunciaremos su contenido de la siguiente manera:

(Ext-Boghossian) Los contenidos mentales y los conceptos involucrados en ellos se determinan, en parte, por aquellos factores externos que constituyen las causas típicas de nuestras percepciones y preferencias.

(Ext-Moya) Los contenidos mentales y los conceptos involucrados en ellos se determinan, en parte, por aquellos factores externos que han sido empleados como paradigmas en nuestro aprendizaje de los usos lingüísticos propios de nuestra comunidad de hablantes (y, por tanto, en la adquisición de los conceptos relevantes).

De acuerdo con Moya, la plausibilidad de una comprensión causal del externismo como la que pone en juego Boghossian es muy escasa. Supongamos, por ejemplo, que Jorge, el dueño de un valioso diamante, es víctima de un robo inadvertido, de manera que el ladrón coloca en el lugar de la joya una pieza de circonita sin valor, completamente indistinguible del diamante por lo que a Jorge respecta. Si la circonita se convirtiera a partir de ahora en la causa típica de las creencias que Jorge expresa al proferir “el diamante me costó un millón de euros” o “estoy muy orgulloso de poseer un diamante tan valioso”, entonces el partidario del externismo debería sostener, según Boghossian, que estas creencias son falsas, en la medida en que se trata de creencias acerca de la circonita. Que este diagnóstico resulte altamente contraintuitivo no constituye en absoluto una razón para rechazar el externismo. Por el contrario, esto sólo sirve, de acuerdo con Moya, para hacer evidente que la comprensión del externismo que Boghossian pone en juego para plantear sus argumentos resulta de lo más insuficiente.

La alternativa que nos propone este autor, cuyo contenido ha quedado recogido en (Ext-Moya), consiste en individuar nuestros contenidos mentales tomando en consideración, en parte, la naturaleza de los factores externos que han figurado en calidad de paradigmas en el aprendizaje del uso de nuestros términos. De esta manera,

al tener en cuenta que los diamantes (y no las porciones de circonita) fueron las instancias del género natural mediante el que se enseñó a Jorge el uso del término “diamante”, entonces el hecho de que las porciones de circonita sean los aspectos externos que a partir de cierto momento constituyen la causa típica de las preferencias de Jorge en las que figura el término “diamante” no compromete al externista con la afirmación de que dichas preferencias se refieren a la circonita y, lo que es más interesante, no se violentan con ello las intenciones de Jorge, quien pretende estar hablando de los diamantes cuando hace uso de ese término.

Dicho esto, Moya plantea un caso de traslado “realista”, mediante el cual intenta demostrar que su comprensión del externismo permite preservar la capacidad de la víctima del transporte para acceder de manera privilegiada a sus propios pensamientos. Para ello, el autor nos invita a considerar una situación hipotética en la que Jorge se traslada a una comunidad lingüística donde la extensión de los términos “diamante” y “circonita” se halla permutada.¹⁸² El dictamen de Boghossian ante esta situación será, previsiblemente, que si Jorge permaneciera en su nuevo entorno el tiempo necesario para establecer relaciones estables con él, aquellas de sus preferencias en las que figura el término “diamante” pasarían a expresar pensamientos sobre la circonita y, correspondientemente, las preferencias que involucran el término “circonita” expresarían ahora pensamientos sobre los diamantes.

Sin embargo, nos asisten dos razones, según Moya, para no dar este paso. La primera de ellas consiste en que ha sido el diamante (y no la circonita) el género natural mediante el cual se le ha enseñado a Jorge el uso del término “diamante”. Esto es, su variante del externismo garantiza una estabilidad en nuestros conceptos al situar sus factores de individuación primordialmente en el entorno en el que se ha llevado a cabo el aprendizaje de los términos que usamos para expresarlos. Si Jorge ha aprendido el uso del término “diamante” en el entorno del que procede, entonces su traslado a un nuevo entorno no es suficiente para decretar una variación en los contenidos expresados. La segunda razón para evitar la conclusión incompatibilista resulta, a nuestro modo de ver, su propuesta más sugerente. Este autor nos dice que “no se requiere que la deferencia [...] sea incondicional: un sujeto puede deferir en una comunidad diferente

¹⁸² Este caso de traslado difiere del contemplado por Boghossian (que involucra a los protagonistas del experimento mental de la Tierra Gemela) en que las comunidades de origen y de acogida de Jorge contienen expertos que conocen con detalle las condiciones de aplicación de los términos “diamante” y “circonita” y que serían capaces de diferenciar una piedra de la otra. Como veremos, esta diferencia tiene una importancia crucial para nosotros.

[de la comunidad en la que se encuentra en la actualidad]” (Moya 2003, p. 181). Esto significa que el mero hecho de que Jorge se traslade a una comunidad que fija unos usos lingüísticos distintos para los mismos términos y, consiguientemente, que expresa conceptos distintos haciendo uso de las mismas expresiones, no compromete indefectiblemente al externista con un cambio en el repertorio conceptual de este hablante. En la medida en que Jorge siga confiando la determinación de la referencia de sus expresiones lingüísticas a los expertos de la comunidad de la que procede, se bloquea la clase de cambios conceptuales inadvertidos que los incompatibilistas pretenden establecer. Uniendo estas dos conclusiones, Moya defiende que la comunidad de deferencia lingüística no es, primordialmente, la comunidad de acogida, sino aquella en la que se lleva a cabo el aprendizaje lingüístico. Según sus propias palabras, “el contenido no queda fijado con completa independencia de las intenciones del sujeto de permanecer fiel a ciertas prácticas y definiciones que determinan el contenido” (Moya 2003, p. 184).

En nuestra opinión, la propuesta de Moya es fundamentalmente correcta, aunque es necesario realizar sobre ella ciertas matizaciones y reforzar con ello su propia plausibilidad. En primer lugar, podría dar la impresión de que su comprensión del externismo presupone que el proceso de aprendizaje de los usos lingüísticos está perfectamente acotado en el tiempo. Si así fuera, aquellos factores naturales que hayan servido de paradigmas en ese proceso determinarían de ahora en adelante y para siempre qué conceptos expresa un hablante haciendo uso de sus expresiones lingüísticas. Sin embargo, aunque esta idea permite bloquear el establecimiento de conclusiones incompatibilistas a partir de casos de traslado (un cambio de entorno en un momento posterior al proceso de aprendizaje *nunca* conllevaría un cambio conceptual), nos comprometería con una concepción muy inverosímil del aprendizaje lingüístico. Por el contrario, cabe suponer que el proceso de aprendizaje se prolonga a lo largo de la vida de un hablante, aunque existan periodos donde los esfuerzos invertidos en ese sentido son especialmente intensos (en la infancia, por ejemplo). No obstante, Moya no ha pasado por alto esta consideración y su respuesta ante posibles objeciones en este sentido consiste en apelar a la formación de conceptos amalgama o conceptos de extensión disyuntiva (2003, p. 184). En el caso de un sujeto que delegue la tarea de determinar la referencia de sus términos en dos comunidades por igual es natural pensar que dicho término tendría un uso mixto, determinado por las condiciones lingüísticas tanto del entorno de origen como del entorno de acogida. Como contrapartida, el

concepto que este hablante expresaría haciendo uso del término tendrá, también, un carácter mixto.

En segundo lugar, aunque la comprensión del externismo que pone en juego la tesis (Ext-Moya) está basada, según este autor, en las intuiciones que respaldan el experimento mental en torno al uso del término “agua”, nos parece sumamente significativo que al plantear un caso de traslado “realista” (usando sus propias palabras) este autor haya escogido un par de términos como “circonita” y “diamante”. Al contrario de lo que ocurre en relación con el término “agua” en el experimento de Putnam, existe un conjunto de expertos relevantes en el uso de los términos “circonita” y “diamante” tanto en el entorno de origen de Jorge como en su entorno de acogida. Si destacamos este aspecto de su trabajo se debe a que pensamos que la plausibilidad de su solución para preservar el acceso privilegiado de los hablantes a sus pensamientos, que pasa por respetar su intención de delegar la determinación del uso de sus términos en una comunidad distinta de la comunidad de su interlocutor, tiene pleno sentido sólo si planteamos un caso de traslado donde exista una diferencia efectiva en el uso que las distintas comunidades establecen con respecto a los términos involucrados. Como se vio, esto no ocurre en el caso del término “agua”, ya que no existe tal diferencia en un entorno y otro ni, correspondientemente, existen expertos que puedan conocerla. De este modo, si planteáramos un caso de traslado tipo como los que formula Boghossian, resultaría indiferente que el hablante que sufre el traslado delegara la determinación de la referencia del término “agua” en una comunidad lingüística distinta de la comunidad de acogida, pues, como ya vimos, las prácticas que en ambos entornos se asocian con ese término son exactamente las mismas. Por consiguiente, resulta cuestionable que la solución de Moya tenga efectividad en un caso de traslado tipo.

No obstante, nuestro veredicto sobre las propuestas de este autor es muy similar al que merece el trabajo de Grimaltós. En las aportaciones de ambos hay algo valioso y básicamente correcto y, si ellas no sirven para solucionar los casos de traslado tipo y la comprensión del externismo que los sustenta, entonces *es esa comprensión del externismo la que debe ser cuestionada*. Lo que proponemos es respetar la accesibilidad privilegiada de primera persona por la vía de respetar las intenciones comunicativas de los hablantes, tal y como estos autores proponen. La dicotomía entre contenidos (C-M) y contenidos (C-P) que estudia Grimaltós y el carácter condicional de la deferencia lingüística defendido por Moya permiten ofrecer el siguiente diagnóstico de los casos de traslado. Lo que se produce en estas situaciones es un *conflicto en la interpretación*,

controlado, localizado y muy semejante al que se manifiesta en el caso de las ambigüedades dialectales. El contenido mental del hablante viene fijado, generalmente, por las convenciones lingüísticas de su comunidad de origen, la cual constituye, tal y como sugiere Moya, la comunidad de deferencia primaria. Si esto es así, el sujeto conocerá sus propias creencias y su intención de expresar el contenido que tienen y no otro distinto, aunque la expresión que ha escogido reciba una interpretación distinta en la nueva comunidad. Adoptar esta solución como el diagnóstico general de los casos de traslado requiere enmarcar el externismo en el ámbito de las prácticas de la interpretación y, con ello, rechazar aquellas variantes del mismo que no sólo resultan irrelevantes para la aclaración de estas prácticas, sino que resultan especialmente agresivas con el acceso privilegiado.

Abogamos, en suma, por comprender el externismo como la tesis que defiende que la individuación de nuestros conceptos requiere tomar en consideración con qué clase de entidades (naturales o no) entramos en contacto, siempre que la identificación de esa clase como tal (y su diferencia con otras clases) esté dentro de nuestro alcance epistémico y/o del alcance epistémico de nuestros interlocutores. Desde luego, una tesis sobre la individuación conceptual no podría denominarse “externista” si no sostuviera que, en ocasiones, algunos factores de individuación de los conceptos de un hablante se encuentran fuera de su cabeza, esto es, fuera de su alcance epistémico individual. Desde luego, respaldamos este punto con nuestra propuesta, pero, a pesar de que un sujeto individual podría poseer un concepto cuyas condiciones de aplicación desconoce en parte, no aceptamos que ese hablante pueda poseer un concepto cuyas condiciones de aplicación son parcialmente desconocidas por todos sus interlocutores. O, dicho de otra manera, rechazamos una comprensión del externismo (*estrictamente natural*), que establezca una diferencia en los conceptos que usan los hablantes en función de las distintas extensiones de sus términos a pesar de que esa diferencia en las extensiones (i) sea desconocida por todos los miembros de su entorno, (ii) sea desconocida por todos los miembros de otros entornos con los que pueden entrar en contacto, (iii) sea una diferencia que, a pesar de ser desconocida, no impide que los hablantes que usan esos términos resulten perfectamente competentes en su uso. La comprensión del externismo que proponemos y que caracterizaremos más precisamente en el último capítulo de nuestro trabajo conlleva, en último término, subordinar la individuación conceptual en función de factores naturales – concebidos como Putnam los concibe – a la individuación externista en función de los factores sociales. Lo que requerimos para

establecer una distinción conceptual efectiva en relación a un mismo término es que dicha distinción tenga efectos semánticos dignos de ser tomados en consideración en los procesos interpretativos que llevan a cabo sus usuarios, en el sentido de que exista una diferencia en las prácticas lingüísticas que regulan su uso.

La consecuencia inmediata de adherirnos a una comprensión del externismo como la que proponemos es que el poder de convicción del experimento mental de la Tierra Gemela en torno al uso del término “agua” desarrollado en las coordenadas originales (año 1750) se verá muy reducido.¹⁸³ El motivo es que este experimento, tal y como es usualmente formulado por los externistas atendiendo a las sugerencias de Putnam en (1975c), pone en juego la variante del externismo de la que nosotros nos desvinculamos. En nuestra opinión, resulta sumamente cuestionable que las diferencias entre los líquidos que reciben el nombre de “agua” en la Tierra y en la Tierra Gemela sean diferencias con los efectos semánticos que exigimos para establecer una distinción en los conceptos que los habitantes de esos entornos expresan mediante dicho término.

En la última sección de este capítulo abordaremos una pregunta de gran importancia en relación a los casos de traslado, a saber, hasta qué punto las situaciones que estos argumentos ponen en juego son alternativas relevantes. Además, profundizaremos en nuestro intento de reducir la mayoría de los casos de traslado que podrían afectar al tipo de externismo que defendemos a un conflicto entre la forma en que el hablante pretende ser interpretado y la manera en que es interpretado por el oyente.¹⁸⁴ Dado que esta clase de conflictos no ponen en peligro (sino que presuponen) el acceso privilegiado de un sujeto a sus propios pensamientos, parece que los casos de traslado no constituyen un problema insuperable para esta concepción del externismo.

¹⁸³ Esto no ocurrirá con los experimentos externistas que toman pie en los ejemplos de Putnam sobre los pares de expresiones “aluminio”-“molibdeno” y “olmo”-“haya”. El motivo es que en este caso la distinción entre las clases de entidades designadas por estos términos sí es relevante para los individuos que reciben la interpretación y para sus interlocutores, algo que va indisolublemente ligado a la división social del trabajo lingüístico.

¹⁸⁴ Reiteramos la idea de que el hablante tiene *derecho* a ser interpretado así, porque hay un marco de normatividad lingüística que respalda esa interpretación. Sin ese marco y si las pretensiones del hablante dominaran por completo el proceso interpretativo éste podría llegar a ser impracticable. Un hablante que se presente a sí mismo como hablante del castellano y que diga de continuo cosas como “tengo un hierro de madera” para referirse a una silla o “las ballenas son dulces” para hablar de los buñuelos no tiene derecho a *imponer* a su interlocutor sus peculiares usos lingüísticos con la excusa de que “eso es lo que pretende decir”.

3. *La relevancia de los casos de traslado*

Cuando en el debate sobre la compatibilidad entre el externismo y el acceso privilegiado surge la cuestión de la relevancia de los casos de traslado lo que se busca es determinar con cuánta frecuencia encontramos situaciones en las que un hablante de un lenguaje natural como el castellano o el inglés podría verse en una situación como la que dramatizan estos casos. Esta pregunta es muy importante puesto que si se demostrara que los casos de traslado son situaciones ficticias sin ninguna influencia en la vida real del hablante medio, entonces el externista podría rehusar hacerse cargo de esta clase de argumentaciones alegando que sus intereses se ciñen sólo a la aclaración de problemas que pueden surgir en el ámbito de las atribuciones mentales ordinarias en el que los casos de traslado tipo son una posibilidad carente de relevancia.¹⁸⁵

La plausibilidad de esta opción se ve reforzada por el hecho de que los casos de traslado tipo apelan a traslados fantásticos entre mundos posibles, algo que, al menos en principio, parece poco factible. No obstante, debemos ir más allá de la apariencia de ciencia ficción que presentan tales casos y determinar si el problema filosófico que estos casos pretenden escenificar es un problema relevante para el hablante real. El esquema de dicho problema puede establecerse así. Lo que se busca es una situación en la que un sujeto:

- (1) Cambia de entorno.
- (2) Permanece en su nuevo entorno un tiempo suficiente para establecer relaciones estables con las nuevas características naturales y/o sociales de éste.
- (3) El sujeto debe ignorar, si no el traslado mismo, al menos el hecho de que su entorno de acogida varía en un sentido relevante.

¹⁸⁵ Desde luego, no han faltado los autores compatibilistas que han ofrecido esta respuesta. Por ejemplo, Warfield en (1992) alega que a menos que se demuestre que los casos de traslado pueden darse en el mundo actual, el argumento incompatibilista no es concluyente. Su idea es que la actualidad de un caso de traslado es condición necesaria para que estos casos dejen de ser meras posibilidades lógicas y pasen a ser alternativas relevantes. Si no se consiguiera demostrar su carácter actual, entonces lo único que el incompatibilista podría concluir es que, dado el externismo, no es el caso que necesariamente un sujeto conoce *a priori* sus pensamientos de manera introspectiva (puesto que habría mundos posibles donde el sujeto carece de esa clase de conocimiento). Ahora bien, negar la necesidad del compatibilismo no demuestra que las tesis externistas sean incompatibles de una manera relevante con el acceso privilegiado a nuestros pensamientos si admitimos, como Warfield propone, que dicho acceso sólo requiere excluir las alternativas relevantes a tales pensamientos (que puedan darse en el mundo actual) y no sólo las alternativas posibles.

Si en el mundo actual puede darse una situación que cumple estas condiciones, habríamos encontrado un caso de traslado relevante. Ludlow en (1995a) intenta argumentar en este sentido invitándonos a tomar en consideración la situación de Biff, un hablante actual del inglés estadounidense, que viaja a Reino Unido, país en el que se habla una variante dialectal de su propia lengua materna. Este caso, como puede verse, es de lo más corriente y no involucra un traslado inconsciente intermundano como en los casos de traslado tipo. Este hablante sabe que hay diferencias semánticas en el inglés de ambas regiones, pero supone que, por lo general, los significados de los términos se solapan. Biff emplea el término “chicory” (endivia) para referirse a cierta verdura que evita comer en la medida de lo posible. Como no es un especialista en botánica, tiene un conocimiento muy básico sobre este alimento y, por ello, delega en la comunidad de botánicos la tarea de determinar las condiciones de aplicación de ese término. Ahora bien, Biff ignora que en Reino Unido el término “chicory” designa la achicoria. Por ello, este hablante supone que al proferir una oración como “I hate chicory” (odio la achicoria/las endivias) está expresando el mismo contenido proposicional en cualquiera de los dos entornos donde reside habitualmente y, por tanto, estará comunicando su odio hacia las endivias. Sin embargo, Biff se equivoca al realizar esta suposición. Así, cuando reside en EE. UU., este hablante expresa mediante dicha oración el mismo contenido que cree estar expresando – que odia las endivias –, debido a que los usos lingüísticos de EE. UU. determinan dicho contenido como un contenido acerca de esa verdura. Pero cuando se traslada a Reino Unido los hábitos lingüísticos comunitarios asociados con el término “chicory” difieren y, por tanto, la preferencia de Biff expresará un contenido distinto de acuerdo con esa asignación semántica. Visto que el examen externista de los pensamientos de este hablante nos invita a establecer una diferencia en sus pensamientos que él mismo es incapaz de advertir, entonces tenemos un caso de traslado en un hablante del mundo actual que cumple las tres condiciones enunciadas en lo anterior.¹⁸⁶

¹⁸⁶ Warfield ofreció una réplica en (1998) a este argumento de Ludlow en la que afirmaba que el incompatibilista debe demostrar que el externismo *implica* falta de acceso privilegiado o, lo que es igual, que en todos los mundos posibles en que el externismo es verdadero la falta de acceso privilegiado también lo es. Sin embargo, según Warfield, Boghossian sólo ha probado que en algunos mundos posibles estas dos tesis son inconsistentes y Ludlow, que son inconsistentes en el mundo posible que es el mundo actual. Lo máximo que estos autores han conseguido establecer es, por tanto, que el externismo es *consistente* con una falta de acceso privilegiado en algunos mundos posibles, pero no que la primera tesis *implica* la segunda. No obstante, las exigencias que Warfield impone al adversario incompatibilista nos parecen un tanto excesivas. Si se consiguiera demostrar que existen tensiones entre el externismo y el acceso privilegiado en el mundo actual ello sería más que suficiente para que las alegaciones incompatibilistas merecieran toda nuestra atención, como hablantes del mundo actual que somos.

La conclusión de Ludlow en (1995a) nos parece válida y pensamos que esta clase de casos de traslado son, tal y como él alega, alternativas relevantes. Sin embargo, es necesario hacer dos observaciones. La primera es que las tesis de este autor no permiten, por sí mismas, establecer el incompatibilismo. Para ello, haría falta demostrar que, dado el externismo, las situaciones relevantes que satisfacen los tres requisitos de los casos de traslado suponen una genuina falta de acceso privilegiado. Por ejemplo, se podría alegar, como vimos, que la capacidad de un hablante para advertir esa clase de cambios conceptuales no entra dentro del alcance y competencia del acceso privilegiado (Falvey y Owens 1994; Goldberg 2003a). En segundo lugar, debe señalarse que Ludlow ha abogado por el carácter actual de los casos de traslado *dado el externismo social*. Esto es, según Ludlow, existen situaciones relevantes en las que un hablante se traslada de una comunidad lingüística a otra sin advertir el cambio, bien porque ignora que su comunidad de acogida establece unos usos lingüísticos distintos para ciertos términos, o bien porque asume equivocadamente que no existe una diferencia en ese sentido (como ocurre en el caso de Biff y el inglés de EE. UU y Reino Unido).

Ahora bien, el hecho de reconocer que los casos de traslado son una alternativa relevante para el externismo social no pone en peligro la accesibilidad privilegiada de un hablante con respecto a sus propios pensamientos. La razón es que la mayor parte de estas situaciones pueden explicarse, según vimos, a partir de un desfase localizado entre el contenido del pensamiento de un sujeto y el contenido semántico de su preferencia, interpretada de acuerdo con los estándares lingüísticos de su comunidad de acogida. De esta forma, cuando Biff dice “I think chicory salad was tasteless” refiriéndose a la ensalada de endivias que cocinaba su madre, lo que este hablante expresa es su creencia de que la ensalada de endivias es insípida. Si interpretamos esta oración en el inglés americano, el contenido de su preferencia y el contenido de su creencia coinciden, pero cuando este hablante se traslada a Reino Unido, las convenciones semánticas asociadas por su comunidad de acogida con el término “chicory” cambian y, debido a esto, su preferencia será interpretada como una preferencia acerca de la achicoria.¹⁸⁷

Sin embargo, deberíamos preguntarnos si este cambio en los nuevos estándares semánticos compromete al externista con un cambio inadvertido en el repertorio conceptual del hablante. En nuestra opinión y siguiendo a Moya, en la mayoría de los

¹⁸⁷ Una ensalada de achicoria puede resultar un plato un tanto extraño. Sin embargo, supongamos, por mor de la argumentación, que los interlocutores británicos de Biff le conceden una gran credibilidad y consideran que la gastronomía estadounidense tiene aspectos extravagantes. Bajo estas suposiciones cobraría mayor verosimilitud la atribución de esa creencia a Biff por parte de sus interlocutores británicos.

casos es muy posible resistirnos a establecer dicha conclusión. Para justificarlo, retomemos el caso de Biff. Este agente es un hablante nativo del inglés americano, relativamente competente en el uso del término “chicory”, que pretende estar hablando de las endivias cuando usa ese término. Es normal, sin embargo, que los hablantes de su entorno de acogida (Reino Unido) atribuyan a este hablante una creencia sobre la achicoria e interpreten sus preferencias como preferencias acerca de este sucedáneo del café, puesto que se encuentran en una comunidad lingüística que respalda dicha interpretación, pero eso no significa que Biff no sepa lo que cree cuando se encuentra en ese entorno. Él piensa que la ensalada de endivias que preparaba su madre era un plato insípido, era *ese* contenido el que pretendía expresar, y, si un hablante le atribuye una creencia distinta, entonces se tratará de una atribución inapropiada. Por esta razón, el conflicto reside en la interpretación de los oyentes y en sus atribuciones, pero no en el conocimiento que este sujeto tiene sobre sus creencias, que permanece intacto.

Es cierto que este caso de traslado, precisamente por su carácter realista, no respeta por completo una de las condiciones de los casos de traslado usuales, a saber, el hecho de que los géneros designados por el término “chicory” son completamente indiscernibles por lo que al protagonista del traslado respecta. Así, supongamos, por ejemplo, que uno de los amigos británicos de Biff le prepara una ensalada de achicoria pensando que se trata de un plato habitual en EE. UU. Si este sujeto viera tal ensalada cabe pensar que se daría cuenta de que *eso* que su interlocutor denomina “chicory salad” no es una ensalada de endivias, dado que las endivias y la achicoria se diferencian a simple vista. Sin embargo, tanto el caso de traslado que propone Moya, que involucra los términos “circonita” y “diamante”, como un caso de traslado planteado en torno al uso de los términos “aluminio” y “molibdeno” en las coordenadas de Putnam en (1975c), serían situaciones mucho más respetuosas con el supuesto de la indiscernibilidad subjetiva. Ahora bien, como hemos explicado, no pensamos que esta clase de casos de traslado planteen un genuino desafío a la compatibilidad entre externismo y acceso privilegiado.

Por otro lado, se podría alegar que nuestra solución (y el aspecto de la solución de Moya en el que nos basamos) no responde de manera apropiada al desafío incompatibilista de los casos de traslado por la siguiente razón. Se diría, según esto, que nosotros esquivamos la conclusión de los casos de traslado impidiendo que estos tengan lugar, esto es, negando que se produzca un cambio conceptual como el que ponen en juego estos argumentos. Si diagnosticamos la situación de Biff en la comunicación con

sus interlocutores británicos como una situación en la que se da un conflicto en la interpretación de sus palabras, es porque no hemos permitido que sus palabras cambien de significado y que pasen a expresar un concepto distinto. En suma, alegar que este hablante se acoge primordialmente a los estándares de su comunidad para poder negociar la interpretación que recibe por parte de los oyentes en caso de que se detecte la divergencia entre su contenido mental y el contenido semántico de la interpretación recibida, es, en suma, bloquear el cambio semántico y conceptual.

Lo primero que hay que decir al respecto es que, como sugiere Moya, de producirse esa clase de cambios, lo harán con menor frecuencia de lo que supone el incompatibilista. Creemos que en la mayoría de los casos un hablante tenderá a apelar al significado de los términos de su lengua materna y a someter la interpretación de sus palabras al arbitrio del marco de normatividad que la respalda. Sin embargo, es posible pensar en la situación de un hablante que viaja a una comunidad lingüística en la que se asigna un significado distinto a uno de sus términos. Para respetar máximamente los supuestos de un caso de traslado, lo mejor es pensar que en dicha comunidad lingüística se habla una variante dialectal de la lengua materna del hablante.¹⁸⁸ Supongamos que la víctima de traslado es semi-competente en el uso de ese término, de manera que sólo conoce *algunas* de las condiciones de aplicación que los expertos relevantes de su comunidad lingüística de origen establecen. Imaginemos, además, que esas condiciones de aplicación del término son *todas* las que determinan los expertos relevantes de su comunidad de acogida. Además, nuestro hablante cree, equivocadamente, que no existe una diferencia en el significado que las dos comunidades lingüísticas asignan al término. El último supuesto que necesitamos para construir este ejemplo es el hecho de que este hablante se encuentra igualmente comprometido con los estándares lingüísticos de ambas comunidades, de manera que renunciamos a la apelación en primera instancia a la comunidad lingüística de origen y al bloqueo en el cambio conceptual, que es la clave de nuestro tratamiento general de los casos de traslado. En este caso, aceptamos que la individuación externista se salda con un cambio conceptual hacia la adquisición de un concepto amalgama o de extensión disyuntiva, puesto que este sujeto delega por igual las condiciones de determinación de la referencia del término en ambas comunidades. Desde luego, viendo que él considera (equivocadamente) que existe una equivalencia

¹⁸⁸ Esto es lo más cerca que podemos estar de la “indiscernibilidad subjetiva” requerida en los casos de traslado al uso. El hecho de que sean variantes dialectales conllevará que el significado de muchos términos en una y otra lengua se solape.

perfecta en el uso del término en ambas comunidades, este cambio conceptual le habrá resultado inadvertido y, por esta razón, se hace necesario reconocer que el externismo que defendemos no preserva en este caso concreto el Principio de Transparencia del Contenido.¹⁸⁹ Sin embargo, la formación de un concepto amalgama es un resultado que no atenta contra las intenciones de este hablante al usar el término relevante, ya que él pretende usarlo como se usa (normativamente) en ambas comunidades.

Imaginemos, para seguir con el ejemplo, que este sujeto regresa a su entorno de origen y expresa sus pensamientos usando de manera inapropiada el término en cuestión (según los estándares de dicho entorno). Este hablante, ¿sabe lo que piensa en ese caso? No hay razón para suponer lo contrario. Él quiere hablar de lo mismo que hablan sus interlocutores (de ambas comunidades) y, dado que ellos se refieren a cosas relativamente distintas entre sí, él querrá hablar de ambas cosas, aunque ignore que se trata de cosas diferentes. Si el concepto que figura en sus creencias es un concepto amalgama podemos dar sentido a esa pretensión.¹⁹⁰ Por otro lado, es de suponer que si su interlocutor del entorno de origen corrigiera la aplicación del término que ha llevado a cabo la víctima de traslado, esta situación merecería la misma lectura que el caso de Bert y su médico: un hablante puede poseer un concepto aunque tenga un dominio imperfecto del mismo. En este caso, la imperfección en el dominio del término se pone de manifiesto en las aplicaciones inapropiadas del mismo que lleva a cabo el hablante de vuelta en su comunidad de origen. De la misma manera que Bert consideraba, equivocadamente, que las normas que regulan el uso del término “artritis” permiten su aplicación a ciertos dolores musculares, nuestro hablante considera, equivocadamente, que el término relevante se aplica sólo a una familia de fenómenos. De cualquier modo, al igual que en el caso del término “chicory”, en esta situación se produciría un conflicto localizado en la interpretación *resoluble en principio por medios al alcance* de los participantes en la comunicación. Para ello, bastaría con que el hablante consultara un

¹⁸⁹ Moya afirma que su solución a los casos de traslado respeta ese Principio, pero no creemos que sea cierto. La razón es que, según él, un caso de traslado puede conllevar la transformación de un concepto unívoco en un concepto amalgama (Moya 2003, p. 184). Ahora bien, es de suponer que dicha transformación le habrá pasado inadvertida al propio individuo, dado lo cual sus contenidos mentales no le resultan completamente transparentes.

¹⁹⁰ A consecuencia de esto, la preferencia en la que figura el término relevante mal aplicado según los estándares de su comunidad de origen será una preferencia falsa en dicha comunidad. Sin embargo, el hablante sabe lo que cree, tendrá una creencia verdadera, aunque haya escogido una expresión para expresar su pensamiento que no es del todo adecuada para que la interpretación del oyente pueda capturarlo. Al igual que antes, la interpretación del oyente es legítima (según sus estándares), aunque en este caso no capture *completamente* lo que el hablante quiere decir. En este último caso, lo que el hablante pretende expresar no es un capricho o una ocurrencia (hablante Humpty Dumpty), sino que está sujeto a dos marcos de normatividad que entran en cierto conflicto en el caso de este término particular.

diccionario o tuviera noticia del equívoco por medio de un hablante que conoce la diferencia existente en el uso del término que establecen ambos entornos.

Desde nuestro punto de vista, la mejor solución para cualquiera que considere que existen intuiciones que respaldan el externismo es encontrar una comprensión del mismo que haga de la individuación conceptual una tarea al alcance de los usuarios del lenguaje. Estimamos que el mejor modo de lograr esto es insertar el externismo en el marco de nuestras prácticas ordinarias de interpretación. En dicho marco, las tesis externistas constituirían la teoría de la individuación más apropiada para responder al fenómeno de la deferencia lingüística. Puesto que los participantes en las prácticas de interpretación tienen la intención de adecuarse a unos estándares lingüísticos que, con frecuencia, ellos mismos no dominan por completo, pensamos que es posible utilizar este dato para encontrar un modo de respetar el acceso privilegiado por medio del respeto a las intenciones de los hablantes. La adhesión a unos estándares comunitarios por parte del sujeto es el factor responsable de la estabilidad de sus conceptos y de la posibilidad de corregir, negociar y refinar su uso. Justamente, el fenómeno de la rectificación y de la negociación en las interpretaciones recibidas constituye la clave para tratar las situaciones que se generan en los casos de traslado realistas en tanto que conflictos en la interpretación.

4. *Conclusiones*

En este capítulo hemos examinado dos variantes del argumento incompatible basado en los casos de traslado. La primera versión, formulada por Boghossian en (1989), defiende que si los contenidos mentales de la víctima de un traslado inadvertido reciben una individuación externista, después del traslado este hablante será incapaz de recordar algunos de los contenidos mentales que poseía en su entorno de origen.

Las réplicas a este argumento que ofrecieron los partidarios del externismo son, como pudimos apreciar, difícilmente conjugables entre sí. Brueckner, por ejemplo, apelaba a una epistemología de corte fiabilista para afirmar que en un caso de traslado no se dan las condiciones necesarias para que el sujeto pueda conocer sus pensamientos pasados, a pesar de que dicho sujeto no haya olvidado nada. Ludlow, por su parte, proponía hacer uso de una “explicación externista de la memoria”, que lleva a concluir que la individuación de los contenidos de los pensamientos rememorados depende de las condiciones externas del entorno en el que se lleva a cabo la rememoración. Esta

solución, como indicamos, no parece demasiado apropiada para mantener la compatibilidad entre el acceso privilegiado y el externismo ya que, según parece, se desprende de ella que la víctima de traslado falsea sistemáticamente sus recuerdos mientras permanezca en su nuevo entorno. Por esta razón, en (2004) Ludlow complementa esta propuesta con una segunda sugerencia que consiste en rechazar que nuestros pensamientos tengan un único contenido. De este modo, este autor intenta mantener que podemos tomar nuestros pensamientos como matrices de contenidos, dado lo cual los diferentes contenidos que una víctima de traslado puede pensar o recordar podrán ser contenidos que pertenecen, todos ellos, a un mismo pensamiento. Sin embargo, argumentamos en detalle que existen serias razones para dudar del carácter genuinamente externista de esta propuesta. Tye, por otro lado, es un autor compatibilista que acepta la tesis de Boghossian de que la individuación externista de los contenidos conlleva, en caso de producirse un cambio de entorno, una pérdida de los antiguos conceptos, admitiendo que dicha pérdida puede asimilarse a un fallo de la memoria. Por último, Burge y Gibbons sostienen que el traslado implica una variación en el repertorio conceptual de la víctima que no tiene por qué traducirse en una pérdida de los antiguos conceptos. De este modo, si el sujeto no carece en su nuevo entorno de los conceptos y contenidos adquiridos en su entorno de origen, entonces el recuerdo no es de principio imposible. Burge, además, postula la existencia de una facultad, denominada “memoria preservativa”, que se encargaría de recordar contenidos adquiridos en el pasado *conservando* su identidad aunque no otorgue al sujeto un *conocimiento* de dicha identidad ni tampoco la posibilidad de diferenciar ese contenido de otros contenidos presentes.

La segunda versión de los casos de traslado, desarrollada por Boghossian en (1992) y (1994), pretende mostrar que un partidario del externismo que mantenga una concepción del acceso privilegiado incompatible con la capacidad de un agente para discriminar entre sus pensamientos debe afrontar consecuencias contraintuitivas. En concreto, la incapacidad de este hablante para detectar la diferencia que el externismo establece entre aquellos contenidos que expresa mediante preferencias homófonas le lleva a cometer falacias de equivocación en sus razonamientos, impropias de un sujeto caritativamente racional.

En este segundo caso, las réplicas de los autores compatibilistas presentaban mayores puntos de coincidencia, en la medida en que todas ellas apelan en mayor o menor medida a las intenciones de los hablantes a la hora de emplear sus contenidos

mentales en calidad de premisas de un razonamiento. Burge, por ejemplo, vuelve a hacer uso de la memoria preservativa para fundamentar la intención caritativa de univocidad conceptual. En lo que respecta a Grimaltós, este autor establecía una distinción entre usos deferenciales y referenciales de nuestras expresiones, en función de si los contenidos pensados por un agente coinciden (uso deferencial) o no (uso referencial) con los contenidos semánticamente expresados por sus preferencias. En un caso de traslado, las preferencias relevantes de la víctima reciben una interpretación distinta de aquella que recibirían en el entorno de origen. Sin embargo, esto no supone *per se* que los contenidos que él pretende estar expresando mediante dichas preferencias también varíen en el mismo sentido (algo que atentaría contra su conocimiento de primera persona). Según Grimaltós, este hablante estaría refiriéndose mentalmente a un conjunto de entidades que es distinto de la referencia semántica de las expresiones que emplea para comunicar su pensamiento. En cuanto a Moya, este autor denuncia, por un lado, que la comprensión del externismo que sustenta el argumento de Boghossian es excesivamente simplista y, por otro lado, que los casos de traslado que tradicionalmente han sido objeto de discusión tienen un carácter poco realista. Por esta razón, él mismo elabora una concepción de las tesis externistas más sofisticada que tiene como objeto no violentar las intenciones de los hablantes al usar sus términos. Además, según este autor, un traslado con carácter realista no supondría un cambio automático en el repertorio conceptual de un hablante. Por defecto, el hablante confiará la determinación de la referencia de sus términos a su comunidad lingüística de origen, dado lo cual podrá saber lo que piensa y lo que pretende expresar mediante sus preferencias, a saber, aquello que cualquier oyente lingüísticamente competente de su entorno interpretaría atendiendo a los usos comunitariamente asociados a los términos que figuran en estas preferencias.

Según se dijo, consideramos que las propuestas de Grimaltós y Moya se encuentran bien encaminadas en el sentido de que ambas escogen el respeto a las intenciones de un hablante como el mejor modo de no atentar contra su privilegio epistémico de primera persona. Aún más, esta elección nos permitiría entender el acceso privilegiado a los propios pensamientos sin caer en la denostada concepción perceptiva de la introspección. Como vimos, ambos autores comparan la situación de la víctima de un caso de traslado con la situación de un hablante que pretende decir algo distinto de lo que su oyente interpreta. Los *lapsus*, el uso de falsos amigos, la asignación de diferentes significados a un término según variantes dialectales de un mismo idioma constituyen,

argüimos, fenómenos donde se presenta de manera clara y controlada esa dicotomía entre el contenido semánticamente expresado por una oración asertiva y el contenido que el hablante pretende comunicar mediante dicha oración. Según creemos, el intento de contemplar la mayoría de los casos de traslado como conflictos en la interpretación requiere desarrollar una concepción del externismo que proporcione a la víctima del traslado el mismo recurso que tienen los hablantes cuyas preferencias resultan inapropiadas para expresar sus pensamientos en las tres situaciones examinadas, a saber, la posibilidad de rectificar y/o negociar la interpretación que recibe.

El primer paso para lograr nuestro objetivo consiste en rechazar aquella concepción del externismo que hemos denominado “externismo estrictamente natural”, derivada del experimento mental de Putnam en torno al uso del término “agua”. Tal y como lo vemos, si las condiciones de uso de un término como “agua” son idénticas tanto en la Tierra como en la Tierra Gemela para *todos* los hablantes de esas comunidades, no encontramos razón alguna para establecer una diferencia semántica en los usos de ese término que llevan a cabo ambas comunidades lingüísticas, ni tampoco en el concepto que sus habitantes expresan al proferir oraciones que lo involucran. Dicho de otra manera, no es descartable que el término “agua” tenga una extensión disyuntiva en ese escenario y el motivo es que las dos alternativas habituales a esta tesis nos parecen igualmente implausibles. Por un lado, se podría defender que el término “agua” tiene un significado distinto según sea usado en un entorno u otro, aunque ninguno de los usuarios del castellano terrestre y del castellano gemelo terrestre lo sepa. Sin embargo, esto nos compromete con una idea tan poco intuitiva como la de que sólo un “oyente póstumo” o “venido del futuro” es capaz de detectar diferencias en el significado de los términos que usan los hablantes de un lenguaje en un momento dado (Castañeda 1989, p. 111). Por otro lado, podría mantenerse que, aunque no existen diferencias en el significado de ese término cuando es usado bien por Óscar₁, bien por Óscar₂, estos hablantes expresan diferentes conceptos cuando emplean el término “agua”. Pero afirmar tal cosa contravendría nuestras prácticas habituales de interpretación y adscripción de contenidos, basadas en el supuesto pre-analítico de que la identidad o diferencia de significados supone una identidad o diferencia de contenidos correspondientes.

Una vez explicados los motivos para nuestro rechazo de esta comprensión del externismo, resulta conveniente hacer una precisión. Pensamos que la versión por la que abogamos no sólo respeta las pretensiones comunicativas de los hablantes (y, con ellas,

su acceso privilegiado) en los casos de traslado, sino que dicho respeto se encuentra en la base de aquellas intuiciones que la respaldan. De esta manera, apelar a las intenciones del hablante no constituiría un recurso *ad hoc* para resolver los problemas del externismo con respecto al acceso privilegiado, sino que esta tesis, en nuestra opinión, constituye el mejor recurso en términos conceptuales para dar cuenta de las intenciones de los hablantes en otras muchas situaciones. En particular, creemos que la lectura externista de los casos de dominio conceptual incompleto, como el que pone en juego el experimento mental de Burge, es la alternativa más caritativa con las intenciones del hablante (Jackman 2000). El dato fundamental de esta clase de situaciones es el hecho de que el hablante que capta un concepto de manera incompleta y lleva a cabo una mala aplicación del mismo reconoce la corrección de su interlocutor como una rectificación del mal uso de *ese* concepto.

La variante de externismo que defenderemos en el último capítulo se apoya fuertemente en el fenómeno de la deferencia lingüística y se inscribe en el marco de las prácticas de interpretación y atribución ordinaria de contenidos. Son estos hábitos que rigen las prácticas de interpretación los que nos han proporcionado, como vimos, las pistas para enfrentarnos a los casos de traslado. Si tenemos en cuenta, además, que la gran mayoría de los traslados relevantes son situaciones en las que un hablante y un oyente se rigen por distintos estándares lingüísticos, estos casos podrían diagnosticarse como una ruptura localizada entre el contenido mental del hablante y el contenido semántico asignado a sus preferencias según los estándares del intérprete. Aunque el hablante no conozca esa diferencia entre los estándares lingüísticos en el momento de ser interpretado, en principio está dentro su alcance y/o del alcance de sus interlocutores detectar dicha ruptura y acogerse a una rectificación y/o negociación. Un conflicto como éste sólo se plantearía, en cualquier caso, sobre la presuposición de que el hablante sabe de manera privilegiada qué es lo que quería decir.

Capítulo 4: Externismo y acceso *a priori* al mundo externo

En los capítulos precedentes hemos examinado dos argumentos cuyo objetivo es mostrar que existen dificultades a la hora de conjugar el externismo con el acceso privilegiado a los propios pensamientos. Tanto el planteamiento incompatibilista simple, tratado en el capítulo 2, como los casos de traslado, analizados en el capítulo 3, se apoyan en la tesis de que el carácter externo de algunas de las condiciones de individuación de los contenidos de nuestros pensamientos nos impediría *lograr* un conocimiento privilegiado de tales pensamientos.

En el presente capítulo vamos a ocuparnos de una nueva propuesta desarrollada principalmente en tres trabajos, a saber, McKinsey (1991), Brown (1995) y Boghossian (1998). Dicha propuesta, que denominaremos de manera genérica *problema o argumento de la consecuencia*, da forma a las intuiciones del incompatibilismo de un modo peculiar.¹⁹¹ Ya no se alega, como en los casos anteriores, que existen dificultades para lograr el auto-conocimiento, dado el externismo. En vez de eso, se asume provisionalmente la verdad del compatibilismo y se intenta reducir al absurdo esta tesis mostrando que, en caso de conjugar el externismo y la obtención privilegiada de auto-conocimiento, tendríamos que hacer frente a consecuencias altamente contraintuitivas. En concreto, lo que se defiende es que si fuéramos capaces de conocer *a priori* nuestros propios pensamientos, pensamientos cuyo contenido, en ocasiones, depende de factores externos (naturales o sociales) de carácter empírico, entonces tendríamos el poder de conocer, igualmente *a priori*, los condicionantes empíricos que los determinan. Dado que, al menos en principio, resulta poco plausible que una teoría filosófica nos provea de la capacidad de obtener conocimiento *a priori* de cuestiones fácticas, naturales o sociales, entonces sólo queda concluir que el compatibilismo es inviable.

Una vez expuesto el esquema de la propuesta incompatibilista de la consecuencia, conviene hacer tres observaciones. En primer lugar, en todas las formulaciones de esta propuesta se asume que el privilegio del conocimiento en primera persona equivale al carácter *a priori* de dicho conocimiento. En segundo lugar, el argumento de la consecuencia se dirige contra la pretensión de defender simultáneamente el externismo y el acceso privilegiado o, lo que es igual, persigue

¹⁹¹ La clasificación de las propuestas incompatibilistas en problemas del *logro* del auto-conocimiento (planteamiento simple y casos de traslado) y problemas de la *consecuencia* del auto-conocimiento es una aportación de Davies en (1998), p. 340 y (2000), p. 390.

reducir al absurdo la conjunción de estas dos tesis. Por este motivo, la derivación de consecuencias inaceptables a partir del compatibilismo no indica, por sí misma, qué miembro de la conjunción debe ser rechazado.¹⁹² En tercer y último lugar, los defensores del argumento de la consecuencia suponen como auto-evidente que la posibilidad de conocer *a priori* las condiciones empíricas de individuación es una consecuencia no deseable del compatibilismo. Ahora bien, algunos autores han optado por rechazar esta suposición, lo que, de aceptarse, haría que el argumento incompatibilista perdiera toda su fuerza.

El plan que proponemos para abordar todas estas cuestiones es el siguiente. Comenzaremos por examinar cuidadosamente los trabajos de McKinsey, Boghossian y Brown, que, como hemos dicho, sientan las bases del argumento de la consecuencia. A continuación, nos ocuparemos de las respuestas compatibilistas en el siguiente orden. En primer lugar, recogeremos aquellas objeciones que ponen en cuestión o bien la verdad de al menos una de las premisas del argumento o bien su carácter *a priori*. En segundo lugar, analizaremos una iniciativa compatibilista que defiende que el acceso *a priori* a cuestiones fácticas no es en absoluto un resultado implausible. Antes bien, si el compatibilismo fuera capaz de proporcionarnos esa clase de acceso epistémico a cuestiones de hecho, habríamos obtenido un gran hallazgo filosófico que podría ser empleado en argumentaciones anti-escépticas. La tercera estrategia compatibilista que abordaremos acepta, igualmente, que las premisas del argumento de la consecuencia son verdaderas y *a priori*, pero alega que dicho razonamiento es impotente para obtener *a priori* el tipo de conocimiento empírico que figura en su conclusión y, por tanto, resultaría inservible para construir refutaciones anti-escépticas. De ser correcto el diagnóstico de estos autores, el argumento de la consecuencia no permitiría obtener conocimiento (*a priori*) de su conclusión (empírica) y, en definitiva, la alegación incompatibilista quedaría sin fundamento.

1. *Versiones del argumento incompatibilista de la consecuencia*

Esta primera sección está dedicada al examen de los argumentos de McKinsey (1991), Boghossian (1998) y Brown (1995). Según dijimos, los trabajos de estos tres autores

¹⁹² Es más, entre los autores pioneros en la formulación de esta propuesta no existe consenso sobre este aspecto. De este modo, mientras que Boghossian considera que el externismo es la tesis responsable del absurdo, McKinsey propone rechazar (cierta modalidad de) el acceso *a priori* a los propios pensamientos.

tratan de establecer la misma conclusión, a saber, que de mantener conjuntamente el externismo psicológico y el acceso *a priori* a los propios pensamientos podremos acceder, igualmente *a priori*, a las condiciones empíricas, de carácter natural o social, que, según el externismo, determinan los contenidos de nuestros pensamientos. Sin embargo, a pesar de esta comunidad en sus pretensiones es conveniente analizar los tres argumentos por separado, ya que sus interlocutores polémicos así como sus presupuestos varían sensiblemente. Comenzaremos con la denominada “paradoja de McKinsey”, desarrollada por este autor en (1991) y defendida en trabajos posteriores como (2002a), (2002b), (2003) y (2007).¹⁹³ A continuación nos ocuparemos de la *reductio* de Boghossian, un argumento que trata de demostrar que la conjunción del acceso privilegiado y del externismo natural basado en el experimento mental de la Tierra Gemela de Putnam da lugar a consecuencias absurdas y, finalmente, atenderemos a las propuestas de Brown de (1995), similares a las de Boghossian pero dirigidas esta vez contra el externismo social de Burge.

1.1 *La paradoja de McKinsey*

Esta primera versión del argumento de la consecuencia presenta la forma de un trilema y pretende concluir que no es posible sostener simultáneamente, so pena de inconsistencia, las siguientes tesis: (a) el conocimiento *a priori* de los propios pensamientos, (b) la individuación de los contenidos de los pensamientos en función de factores empíricos externos y (c) la imposibilidad de conocer *a priori* aquellos hechos empíricos externos que, según el externismo, determinan el contenido de nuestros pensamientos.

La primera premisa del trilema sólo es, para McKinsey, la instanciación de la tesis del acceso privilegiado a nuestros pensamientos. Esto es, cualquiera que sostenga que un sujeto puede saber lo que piensa “sólo con pensarlo”, sin necesidad de emprender investigaciones empíricas ni de realizar ninguna suposición sobre el mundo externo (McKinsey 1991, p. 9), estará comprometido con la primera premisa del trilema:

¹⁹³ Debemos decir que a lo largo de esta sección vamos a exponer el contenido de la paradoja siguiendo únicamente las pautas de McKinsey en (1991), aun cuando este autor realiza aclaraciones posteriores sobre este argumento en otros artículos como (2002a), (2002b), (2003) y (2007). Preferimos reservar el examen de esos últimos trabajos para el apéndice a este capítulo porque las propuestas de McKinsey verdaderamente influyentes en el debate sobre la compatibilidad entre el externismo y el acceso privilegiado han sido sólo las expuestas en el artículo de (1991).

(McK 1) Óscar₁ sabe *a priori* que está pensando que el agua es húmeda.

En cuanto a la segunda premisa, McKinsey considera que las tesis del externismo psicológico natural y social establecen que los contenidos de los pensamientos que adscribimos a un sujeto *dependen necesariamente* de condiciones externas (naturales o sociales) de carácter empírico (McKinsey 1991, pp. 9-12). Desde luego, cabría alegar que esta descripción de las tesis externistas resulta un tanto crasa, ya que los partidarios del externismo más reconocidos no estarían de acuerdo en que *todos* los pensamientos que adscribimos a un sujeto dependen necesariamente de ciertas condiciones externas, ni tampoco suscribirían la tesis de que la identidad de los contenidos de nuestros pensamientos depende *sólo* de condiciones de esa clase, dos matizaciones que McKinsey no recoge en su definición. No obstante, aceptaremos la caracterización del externismo que este autor nos propone con el fin de proseguir en la exposición de la paradoja. Para ello sería conveniente indagar en torno a la naturaleza de la relación de dependencia necesaria que el externismo establece entre los contenidos de nuestros pensamientos y las condiciones sociales o naturales del mundo externo, previamente a lo cual distinguiremos, con McKinsey, entre la relación de implicación/dependencia conceptual, por un lado, y la relación de implicación/dependencia metafísica, por otro. Seguidamente, tomaremos en consideración el argumento que él formula para demostrar que la relación que el externismo establece entre nuestros pensamientos y los factores empíricos que condicionan sus contenidos es una relación conceptual o lógica y no metafísica. Con el fin de analizar su razonamiento, comencemos por recoger su caracterización de la *implicación conceptual o lógica*:

“Digamos que una proposición *p* *implica conceptualmente* [o lógicamente] una proposición *q* [y que, correspondientemente, *q* depende conceptualmente de *p*] si y sólo si hay una deducción correcta de *q* desde *p*, una deducción cuyas únicas premisas distintas de *p* son verdades necesarias o conceptuales que son cognoscibles *a priori*, y cada uno de cuyos pasos se sigue de las líneas previas por una regla de inferencia auto-evidente de algún sistema de deducción natural adecuado” (McKinsey 1991, p. 14).

Por otro lado, en lo que respecta a la relación de *dependencia metafísica*, este autor no ofrece una definición tan explícita. Sin embargo, de sus declaraciones cabe inferir que un individuo (o pensamiento) *S* depende metafísicamente de otro individuo *R* si y sólo si *S* requiere la existencia de *R* para existir en tanto que *S* (McKinsey 1991,

pp. 13 y ss.). McKinsey emplea un ejemplo de Kripke (1980, p. 113) para ilustrar esta clase de relación. Así, la dependencia que existe entre un hablante y sus padres biológicos es de carácter metafísico, ya que en función de que sus progenitores existieron y tuvieron la identidad que tuvieron, un hablante existe y tiene la identidad que tiene.

El razonamiento que despliega McKinsey con el fin de demostrar que la relación de dependencia entre nuestros pensamientos y los condicionantes externos de sus contenidos es de carácter conceptual o lógico y no de carácter metafísico consiste en una reducción al absurdo de esta segunda posibilidad. Su argumento discurre del siguiente modo. Si la relación que el externismo postula entre nuestros pensamientos y los condicionantes externos de sus contenidos fuera de carácter metafísico, entonces el externismo mantendría que nuestros pensamientos son estados psicológicos que *presuponen la existencia* de un individuo distinto de aquél al que se le atribuye el pensamiento. O, lo que es igual, los externistas defenderían que nuestros pensamientos son estados psicológicos en sentido amplio, atendiendo a la definición de dichos estados aportada por Putnam (1975c, pp. 119 y ss.). Ahora bien, dice McKinsey, de acuerdo con la caracterización ofrecida de la relación de dependencia metafísica, esto supondría que *todos* los estados psicológicos de un hablante son estados psicológicos amplios *pero en un sentido trivial*. Dado que el externismo establece que los pensamientos de ese sujeto dependen metafísicamente de factores externos y la existencia de cualquier hablante presupone metafísicamente la existencia de sus padres biológicos, entendidos como individuos distintos del sujeto de atribución, entonces cualquier pensamiento que se le atribuya a este hablante será un estado psicológico en sentido amplio. La razón de esto es que, en cuanto que se trata de un pensamiento de *ese* sujeto, su atribución efectiva depende de la existencia de los progenitores de dicho sujeto, puesto que se requiere que sus padres existan para que él pueda existir *como sujeto de atribución de ese estado*. Pero, como este resultado es absurdo, entonces, según McKinsey, sólo nos queda entender la relación de dependencia que el externismo establece entre nuestros pensamientos y los factores del mundo externo como una relación de dependencia lógica o conceptual.¹⁹⁴

¹⁹⁴ Se comprobará en la próxima sección que este razonamiento de McKinsey ha sido cuestionado por algunos autores compatibilistas. Tampoco a nosotros nos parece demasiado convincente ya que, según argumentaremos, arranca de una alternativa ficticia: o bien el externismo es una tesis sobre la dependencia metafísica de nuestros pensamientos con respecto al entorno externo, entendiendo esa

Después de desplegar esta argumentación McKinsey analiza la dependencia externista que existe con respecto al pensamiento que Óscar₁ expresa mediante la preferencia “el agua es húmeda” y, con ello, ofrece la siguiente instancia de la segunda premisa del trilema:

(McK 2) La proposición de que Óscar₁ está pensando que el agua es húmeda implica conceptualmente E.

En esta instancia (McK 2), E abrevia la condición empírica que, de acuerdo con el externismo, es la responsable de que Óscar₁ esté pensando que el agua es húmeda más bien que cualquier otro contenido.

La tercera premisa del trilema, por su parte, enuncia que no es posible conocer *a priori* aquellos hechos empíricos, de carácter natural o social, de los que, según el externismo, dependen conceptualmente nuestros pensamientos. De esta manera, supongamos que en la instancia (McK 2) “E” abrevia la condición empírica “que hay H₂O en el entorno de Óscar₁”. McKinsey considera obvio que factores de naturaleza empírica como éste no pueden conocerse *a priori* (1991, p. 16), ya que el hecho de que un entorno contenga H₂O (y no XYZ) no es algo que uno pueda averiguar “sólo con pensarlo”, sin realizar ninguna investigación o suposición empírica. Correspondientemente, la instancia de la tercera premisa será la siguiente:

(McK 3) La proposición E no puede ser conocida *a priori*, sino sólo mediante investigación empírica.

En este punto tenemos todos los elementos que se requieren para formular el trilema. Supongamos, en primer lugar, que es posible compatibilizar el acceso *a priori* a nuestros propios pensamientos con el externismo. En este caso, Óscar₁ podría saber *a priori* que está pensando que el agua es húmeda, tal como establece (McK 1) y, asumiendo el externismo instanciado en (McK 2), que Óscar₁ esté pensando que el agua es húmeda depende conceptualmente de E, esto es, del hecho empírico de que su entorno contiene agua. Pero, dado que Óscar₁ sabe *a priori* que está pensando que el agua es húmeda y acepta, como externista, que existe una dependencia conceptual entre el contenido de su pensamiento y E, entonces, atendiendo a la definición de “dependencia conceptual”, este hablante podrá deducir E *a priori* a partir de su pensamiento, sin requerir más que las leyes de la lógica. Esto muestra, según McKinsey, que la conjunción de la aprioricidad del auto-conocimiento en (McK 1) y la verdad del

dependencia *sólo* como él la describe, y entonces el externismo es trivial, o bien sólo nos queda entender el externismo como una tesis sobre la dependencia conceptual.

externismo en (McK 2) es incompatible con la instancia (McK 3), la tesis de que E sólo puede conocerse empíricamente.¹⁹⁵

La paradoja de McKinsey, en suma, pretende establecer la incompatibilidad entre el externismo y el acceso privilegiado mostrando que la conjunción de estas dos tesis nos proporcionaría la inverosímil capacidad de acceder *a priori* a cuestiones de hecho. Aunque en secciones posteriores tendremos oportunidad de examinar con más detalle algunas críticas al argumento de McKinsey, por el momento señalaremos que entre los supuestos más problemáticos del mismo se encuentra su concepción fuerte de la aprioricidad, así como la defensa del carácter conceptual (y no metafísico) de las tesis externistas.

1.2 *La reductio de Boghossian contra el externismo natural*

En esta subsección nos ocuparemos de la versión del problema de la consecuencia propuesta por Boghossian en (1998). Lo primero que debe decirse es que, aunque en ese trabajo se reconoce el parentesco existente entre la nueva versión del argumento que se propone y la paradoja de McKinsey, también figura en él un propósito explícito de proceder con más cautela para remediar ciertas inexactitudes que minan la plausibilidad de dicha paradoja (Boghossian 1998, p. 198, n. 1).¹⁹⁶ De acuerdo con esta voluntad de precisión, Boghossian indica que el objetivo de su argumento es atacar únicamente la versión del externismo psicológico que toma pie en el experimento mental de la Tierra Gemela de Putnam,¹⁹⁷ dado lo cual este autor pondrá máximo cuidado en enunciar los presupuestos de esta tesis y del experimento mental que la sustenta.

¹⁹⁵ Las otras dos alternativas del trilema parecen claras. Si no queremos renunciar a (McK 3) y tampoco al conocimiento *a priori* de nuestros propios pensamientos recogido en (McK 1), entonces debemos rechazar (McK 2), puesto que la dependencia conceptual que el externismo establece entre nuestros pensamientos y E es responsable de que E pueda ser conocido *a priori* si nuestros pensamientos pueden conocerse de la misma manera. Por otro lado, si estuviéramos convencidos de la verdad del externismo, instanciado en (McK 2), y nos pareciera inasumible la posibilidad de conocer E *a priori*, recogida en (McK 3), entonces sólo quedaría renunciar al acceso *a priori* a nuestros propios pensamientos, alternativa por la que optará el propio McKinsey en trabajos posteriores a (1991).

¹⁹⁶ En particular, Boghossian afirma que Brueckner en (1992) ha refutado de manera efectiva la versión del argumento de la consecuencia formulada por McKinsey. En la siguiente sección veremos qué argumento ha empleado Brueckner para este fin.

¹⁹⁷ Hay que decir, sin embargo, que Boghossian señala que aunque en su artículo sólo está tomando en consideración esta versión del externismo “podría darse un argumento paralelo” para la versión social del externismo psicológico (Boghossian 1998 p. 200, n.2). No obstante, Boghossian no ofrece dicho argumento y, como veremos, autores como Brown (1999) han señalado que muchos presupuestos que figuran en Boghossian (1998) no serían directamente aplicables al externismo de Burge.

Boghossian ofrece diversas caracterizaciones de las tesis externistas a lo largo de su trabajo, con grados de generalidad muy distintos. Antes de abordar dichas caracterizaciones, conviene aclarar el significado de algunos de los términos que él usa para enunciar esa tesis y sus supuestos. Boghossian distingue, así, varias propiedades semánticas de nuestros términos de género natural y de las oraciones en las que figuran tales términos.¹⁹⁸ Mientras que este autor define la *extensión* de un término de género natural de una manera bastante usual, es decir, como el conjunto de entidades designado por el término, distingue, por otro lado, la extensión del término de su *referente*, que es la propiedad que denota el término en cuestión. Asimismo, entiende el *significado* de un término de género natural como el *concepto* expresado por el término, aunque rehúsa explícitamente especificar la naturaleza de ese significado y su relación con la extensión y el referente. Por último, la *condición de verdad* de una oración en la que figura un término de género natural es idéntica a la *proposición* que expresa y ésta, a su vez, tiene como componentes las *propiedades* denotadas por los términos de género natural involucrados.

Ofrecidas estas definiciones, Boghossian enuncia la caracterización más general del externismo, entendiendo por tal la tesis de que los hechos externos a un hablante deben ser tomados en consideración a la hora de individuar sus contenidos mentales (Boghossian 1998, p. 199). El externismo natural toma pie, como es bien sabido, en una cierta lectura de la situación descrita en el experimento mental de la Tierra Gemela diseñado por Putnam en (1975c). Esta lectura concluye que la sustancia a la que actualmente se aplica un concepto de género natural – como el concepto de agua – interviene en la individuación del propio concepto (Boghossian 1998, p. 200). No obstante, Boghossian indica que para poder extraer esta conclusión se requieren *dos presupuestos* adicionales, propios de la lectura externista. El primero de ellos es que los líquidos llamados “agua” en la Tierra y en la Tierra Gemela son géneros naturales diferentes, razón por la cual ese término denota una propiedad distinta en cada entorno. El segundo supuesto, que es el que proporciona a esta interpretación su cariz externista, concierne a las propiedades semánticas de ese término. En concreto, el externista sólo

¹⁹⁸ Debe decirse que algunas de estas propiedades semánticas de los términos de género natural y de las oraciones en las que figuran esos términos son propiedades que los términos de género natural comparten con otros términos generales y que también poseen las oraciones en las que no figuran términos de género natural. De este modo, la condición de verdad de una oración (de cualquier oración) es la proposición que expresa dicha oración, si es que expresa alguna. No obstante, en la medida en que sólo nos interesa atender a los términos de género natural y a las oraciones que contienen esta clase de términos, de ahora en adelante aludiremos únicamente a las propiedades semánticas de estos dos tipos de expresiones.

puede extraer una conclusión favorable a su propia posición a partir de la diferencia de referentes si supone, además, que los conceptos expresados por los términos de género natural como “agua” determinan su referente.¹⁹⁹ De aceptar esto, la diferencia que existe en cada entorno entre los referentes del término nos permitiría establecer una diferencia correspondiente en el concepto que expresan los habitantes de estos planetas cuando usan el término “agua”. Boghossian resume esta idea diciendo que para el externismo las condiciones de satisfacción del concepto expresado por un término de género natural y las condiciones de verdad de las oraciones en las que figura uno de estos términos (las proposiciones que cada una de éstas expresan) son, en ambos casos, independientes de contexto. Esta caracterización del externismo es un tanto inusual y estimamos, incluso, que podría dar lugar a confusiones. No obstante, cabe explicar lo que este autor quiere indicar al hablar de independencia de contexto diciendo que, según el externismo, una oración como “el agua es húmeda”, proferida en la Tierra, tiene como condición de verdad determinada que el líquido que posee la propiedad de ser agua también posee la propiedad de la humedad, y no tendrá otra condición de verdad distinta en otro entorno diferente. La independencia de contexto recoge el hecho de que las condiciones de verdad de una oración quedan fijadas por el contexto inicial de su proferencia y conservarán su identidad en (y serán independientes de las peculiaridades de) todos los posibles contextos en los que pueda llevarse a cabo su evaluación semántica.

Tras caracterizar las tesis externistas, Boghossian explicita *tres premisas* que intervienen en la construcción del escenario de la Tierra Gemela (y de la lectura externista que se hace de él) y que le resultarán de utilidad en la formulación de su argumento. La primera de ellas consiste en que el término “agua” *pretende* expresar un concepto de género natural y, además, que la pretensión de usar ese término para nombrar un género natural es algo que los usuarios del lenguaje pueden saber *a priori* (Boghossian 1998, p. 201). Por el contrario, que un término se aplique de manera efectiva a un género natural (o no) es algo que no puede dirimirse sin intervención de la experiencia. Una segunda premisa es que los sujetos que protagonizan el experimento mental son ignorantes en química. Esto conlleva, como vimos, que a Óscar₁ y a Óscar₂ el líquido que denominan “agua” les parece exactamente el mismo por todo lo que

¹⁹⁹ El internista, por su parte, renuncia a esta clase de determinación al mantener que el concepto expresado por el término “agua” (fijado en virtud de factores internos) es el mismo concepto en la Tierra y en la Tierra Gemela, aunque en cada entorno le corresponderá un referente (y una extensión distintas) y, por tanto, las condiciones de verdad de las oraciones en las que figura ese término serán distintas en cada contexto (y dependientes de él en el sentido de Boghossian).

saben *a priori*, ya que uno y otro necesitarían adquirir ciertos conocimientos empíricos para poder distinguirlos. La tercera premisa establece que el término “agua”, así como cualquier otro término de género natural susceptible de recibir una individuación externista, debe expresar un concepto *atómico*. Según el ejemplo que él mismo nos ofrece, si el término “agua” expresara un concepto definido mediante otros conceptos (conceptos como los de líquido inodoro, incoloro e insípido), ese término expresaría un concepto compuesto.

Antes de exponer su argumento y con el fin de hacerlo con la mayor claridad posible, recapitularemos los dos supuestos de la lectura externista del experimento mental de la Tierra Gemela y las tres premisas involucradas en su construcción. Así, (i) el supuesto de que el H₂O y el XYZ son géneros naturales distintos; (ii) el supuesto de que los conceptos expresados por términos de género natural determinan su referente; (iii) la premisa de que los protagonistas del experimento mental son ignorantes en cuestiones químicas; (iv) la premisa de que ambos pretenden expresar un concepto de género natural haciendo uso del término “agua” y, por último, (v) la premisa de que el concepto expresado por el término “agua” es, en ambos casos, un concepto atómico.

Hechas estas aclaraciones, examinaremos el argumento de la consecuencia que propone este autor. En lo que sigue, nos referiremos a esta versión del argumento con el nombre de “*reductio* de Boghossian”, que es la denominación que recibe en la literatura especializada y que refleja su pretensión explícita de reducir al absurdo el externismo compatibilista.

Boghossian establece la primera premisa del argumento como una premisa que un hablante como Óscar₁ podría establecer *a priori*:

(RB 1) Tengo el concepto de agua.

Según Boghossian (1998, p. 198), si un hablante tiene acceso privilegiado a sus propios pensamientos, contenidos y conceptos, entonces es capaz de saber que los posee sin necesidad de emprender investigaciones empíricas o, lo que es igual, puede conocerlos *a priori*. La segunda premisa, por su parte, es un pensamiento que, de nuevo, Óscar₁ podría establecer de manera puramente apriorística a partir de su aceptación del externismo:

(RB 2) Si tengo el concepto de agua, entonces el agua existe.

El establecimiento de esta segunda premisa se justifica porque, según Boghossian, el externismo (una tesis establecida *a priori*) está comprometido con la existencia de los géneros naturales que *pretenden* ser nombrados por aquellos términos

de género natural que expresan conceptos *atómicos*. Esta segunda premisa lleva sobre sí todo el peso del argumento incompatibilista de Boghossian y él mismo es consciente tanto de su centralidad como de su carácter controvertido. Sin embargo, antes de examinar con detalle qué justificaciones nos ofrece para sostener (RB 2), concluiremos la enunciación de la *reductio* mostrando que Óscar₁ sólo requiere emplear un *modus ponens* (una regla de la lógica cognoscible *a priori*) para extraer la siguiente conclusión a partir de (RB 1) y (RB 2):

(RB 3) El agua existe.

De acuerdo con Boghossian, como parece claro que esta conclusión empírica no puede establecerse por medios puramente apriorísticos, entonces sólo nos quedan dos opciones, a saber, o declarar que al menos una de las premisas es falsa o defender que al menos una de ellas no puede ser conocida *a priori*. No obstante, para este autor cualquiera que pretenda respetar la tesis del acceso privilegiado a nuestros propios pensamientos (que equivale para él al acceso *a priori* a los mismos) no puede emplear ninguno de estos dos recursos en relación con la premisa (RB 1), dado lo cual la premisa bajo sospecha será (RB 2). Ahora bien, puesto que la premisa (RB 2) constituye una instancia de una teoría filosófica extraída por medios puramente conceptuales, parece que también debemos asumir su carácter apriorístico. Por esta razón, concluye Boghossian, la única opción que nos queda es reconocer la falsedad del externismo, ya que, como se ha demostrado, su conjunción con el conocimiento *a priori* de los propios pensamientos implica la capacidad de establecer igualmente *a priori* una cuestión fáctica, a saber, la existencia de géneros naturales en nuestro entorno.

Una vez expuesto el contenido de la *reductio* conviene examinar si la premisa (RB 2), que constituye el fulcro de la argumentación incompatibilista de Boghossian, es una tesis que, efectivamente, se desprende del externismo y si, además, se trata de una premisa que puede establecerse *a priori*. Como hemos indicado, el autor es consciente de que (RB 2) será el objeto prioritario de las críticas de los autores incompatibilistas, por lo que invierte un considerable esfuerzo en justificarla.

En primer lugar, Boghossian se adelanta a la objeción compatibilista más natural, a saber, que ningún externista, en tanto que tal, debe requerir la existencia de agua en el entorno de un sujeto para que éste pueda mantener pensamientos sobre ese líquido. Para responder a esta objeción Boghossian se pregunta de qué manera el externismo puede explicar que alguien que vive en un entorno sin agua posea ese concepto y, seguidamente, descarta tres posibles explicaciones. En primer lugar, se podría alegar

que alguien puede tener el concepto de agua aunque no haya agua en su entorno siempre y cuando este sujeto posea un concepto que cumpla el mismo rol funcional que el concepto de agua.²⁰⁰ Pero, si un externista mantuviera que el concepto de agua se individúa por su rol funcional, Boghossian alega que esto sería fatal para su propia posición, puesto que no habría ningún motivo para negar que Óscar₂, el habitante de la Tierra Gemela, también posee el concepto de agua, ya que, por hipótesis, el concepto que expresa haciendo uso del término “agua” tiene el mismo rol funcional que el concepto (de agua) que expresa Óscar₁ haciendo uso de ese término. En segundo lugar, desde el externismo se ha propuesto que un agente puede poseer un determinado concepto de género natural si es capaz de teorizar sobre él, aun cuando dicho género no se encuentre presente en el entorno del agente (Burge 2007, p. 3). No obstante, Boghossian rechaza tajantemente esta solución, puesto que contravendría una de las premisas del experimento mental de la Tierra Gemela, a saber, que Óscar₁ y Óscar₂ son ignorantes en química. En tercer lugar, cabría alegar que un hablante puede poseer el concepto de agua siempre que existan usos lingüísticos comunitariamente asociados al término “agua”. Sin embargo, Boghossian replica que esta propuesta no supondría ningún avance, puesto que podría generarse una *reductio* análoga a la que él propone sustituyendo el consecuente de (RB 2) por una disyunción como “o el agua existe o existen hablantes que han establecido ciertas convenciones lingüísticas para regular el uso del término ‘agua’”. Pero, según este autor, concluir *a priori* que existen otros hablantes (una cuestión de hecho) sería tan absurdo como concluir *a priori* que el agua existe.

Para continuar, Boghossian decide hacer frente a otra posible objeción externista. Así, se diría que aun aceptando la verdad de la premisa (RB 2), cabe afirmar que su verdad no puede establecerse puramente *a priori*, sino que para ello se requieren indagaciones empíricas. El autor responde a esto diciendo que, de acuerdo con las premisas del experimento mental de la Tierra Gemela que él ha enunciado previamente, no se precisan esa clase de indagaciones al respecto. En particular, dicho experimento

²⁰⁰ Putnam, por cierto, parece adherirse a esta segunda opción. Así, “[h]ay palabras que se suponía que eran términos de clase natural pero que no se referían, en última instancia, a especies naturales. Se creyó que ‘flogisto’ era el nombre de una especie natural [...] En estos casos, y por razones obvias, parecería que el factor determinante del significado fuera algo semejante al papel conceptual, pues no sería correcto decir que las palabras ‘flogisto’, ‘éter’ y ‘fluido calórico’ son sinónimos porque tienen la misma extensión (vacía). Al no tener ninguna extensión, es decir, al faltar una extensión no vacía que constituya la individualidad de la palabra, se vuelve, naturalmente a la teoría del papel conceptual. En efecto, esta teoría se acerca mucho a la verdad en el caso de las palabras cuya extensión es vacía” (Putnam 1988, pp. 87-88).

muestra que cualquier término que *pretenda* expresar un concepto *atómico* de género natural ha de recibir una individuación externista. Dado que tanto esa pretensión como la propia atomicidad del concepto son, según él, cognoscibles *a priori*, y, puesto que, además, se ha probado que un hablante sólo puede poseer ese concepto si existe agua en su entorno, entonces Óscar₁ puede deducir *a priori* que existe agua en su entorno.

Sin embargo, Boghossian es consciente de que aún queda una última objeción incompatibilista para impedir el establecimiento de la conclusión de la *reductio*. El externista podría alegar que aquellos conceptos susceptibles de recibir una individuación externista no sólo deben *pretender* designar un género natural, sino que *deben designarlo efectivamente*. En ese sentido, cabe concebir situaciones en las que un hablante podría estar equivocado al pensar que el término “agua” designa de hecho un género natural.²⁰¹ Si además de la *pretensión* de designar un género natural se requiriera la designación efectiva del mismo por parte del término, un agente no podría saber *a priori* si el concepto que expresa mediante el uso de dicho término es un concepto susceptible de recibir una individuación externista y, por tanto, Óscar₁ no podría establecer *a priori* la premisa (RB 2).

A fin de dar respuesta a esta objeción, Boghossian diseña un experimento mental, análogo al experimento mental de la Tierra Gemela, en el que se nos pide tomar en consideración un nuevo entorno, la Tierra Seca, en el que la extensión del término “agua” es vacía. Sin embargo, todos los habitantes de la Tierra Seca son víctimas de una alucinación colectiva que les induce a creer que existe un líquido inodoro, incoloro e insípido que corre por ríos y lagos, sacia la sed de los seres vivos y que se denomina “agua”. A partir de estas premisas Boghossian se pregunta qué clase de concepto expresan (si es que expresan alguno) los habitantes de la Tierra Seca cuando usan el término “agua”. El examen de las dos posibles respuestas (atómico o compuesto) a esta cuestión le servirá al autor para ofrecer un argumento adicional que refuerce su tesis más fuerte, a saber, que la individuación externista de los contenidos de nuestros pensamientos hace que éstos sean dependientes de objeto. Su examen discurre de la siguiente manera:

²⁰¹ Ello ocurriría, según Boghossian, si este hablante estuviera inserto en un entorno en el que ese término designa una mezcla heterogénea de sustancias, al igual que ocurre con el término “jade”, que sirve para designar tanto a la jadeíta ($\text{Na (Al Fe}^{3+}\text{) Si}_2\text{O}_6$) como a la nefrita ($\text{Ca}_2\text{(Mg Fe}^{2+}\text{)}_5\text{(Si}_8\text{O}_{22}\text{)(O H)}_2$). O también podría suceder que el término “agua” no designara ningún género natural.

(A) Supongamos que ese concepto es *compuesto*. Pero esto contravendría el hecho de que el carácter atómico o compuesto de un concepto sobreviene a las propiedades intrínsecas del hablante que lo posee (i.e., está determinado por su sintaxis interna). Dada la estipulación de una identidad intrínseca entre Óscar₁, el habitante de la Tierra, y Óscar₃, un habitante de la Tierra Seca, no puede ocurrir que el primero de ellos esté expresando un concepto atómico y el segundo, un concepto compuesto haciendo uso de ese mismo término.

(B) Supongamos que el concepto es *atómico*. Boghossian se pregunta por el concepto que expresaría Óscar₃ si tuviera un carácter atómico. Su respuesta es que no hay manera de determinar su identidad atendiendo a las pautas externistas de individuación conceptual. Visto que Boghossian establece una ecuación entre proposición y condiciones de verdad y dado que, por hipótesis, la proposición que Óscar₃ expresaría mediante la oración “el agua es húmeda” carece de condiciones de verdad definidas en la Tierra Seca, entonces no parece posible identificar dicha proposición.

Mediante este dilema el autor considera probado que respaldar el externismo supone respaldar una tesis que afirma que, en caso de que un término de género natural carezca de extensión, no habrá ninguna propiedad que constituya el referente del término, ni dicho término expresará concepto alguno ni, correspondientemente, se estará expresando ninguna proposición cuando un hablante profiera una oración en la que figure ese término.²⁰²

El modo en que Boghossian usa este resultado para reforzar la plausibilidad de su *reductio* es el siguiente. Según demuestra el caso de la Tierra Seca, el externista se ve forzado a admitir que los contenidos proposicionales expresados mediante oraciones en las que figuran términos de género natural son dependientes de objeto. Pero, si esto es así, entonces la conclusión de la *reductio* está servida. Cualquiera que pueda saber *a priori* que posee un concepto *atómico* que *pretende* nombrar un género natural, podrá saber, igualmente *a priori*, que ese concepto no es vacío, lo que implica que será capaz de obtener conocimiento *a priori* de un hecho empírico, a saber, que existe la clase de entidades a las que se aplica el concepto relevante.

²⁰² Boghossian es más prudente que McKinsey a la hora de establecer la misma conclusión. No es que el externismo implique lógicamente la dependencia de objeto, sino que el externismo *en conjunción con supuestos adicionales* implica dependencia de objeto (Boghossian 1998, p. 210). Los dos supuestos más importantes en este sentido son los siguientes: (i) los conceptos expresados por términos de género natural de extensión vacía son atómicos y (ii) las propiedades son componentes proposicionales.

Llegados a este punto y tras examinar las propuestas de Boghossian conviene destacar los aspectos más discutidos de la misma, aun cuando reservemos su tratamiento en detalle para la segunda sección. En primer lugar, la concepción de la aprioricidad presupuesta en la enunciación de la *reductio* será objeto de controversia y, en segundo lugar, la distinción teórica entre conceptos atómicos y compuestos se someterá a fuertes cuestionamientos. Como ha podido apreciarse, esta distinción constituye una de las claves para extraer su tesis central, a saber, que los pensamientos que involucran conceptos atómicos que pretenden aplicarse a géneros naturales son pensamientos dependientes de objeto.

1.3 *La reductio de Brown contra el externismo social*

Esta tercera y última parte de la primera sección está dedicada al examen del argumento incompatibilista ideado por Brown en (1995). El núcleo del trabajo de esta autora consiste en ofrecer una versión de dicho argumento centrada no ya en el externismo natural, sino en las propuestas del externismo social de Burge. De esta forma, aunque existe un parecido más que evidente entre la propuesta de Boghossian y la de Brown, esta autora advierte que un defensor del externismo social de Burge posee, en cuanto tal, suficientes recursos como para escapar a la versión del argumento de la consecuencia propuesto por Boghossian (Brown 1999, p. 55). El motivo fundamental es que, dado que esta variante del externismo legitima la atribución de conceptos a un hablante que los comprende sólo parcialmente, cabría pensar que un hablante puede poseer un concepto, saber (*a priori*) que lo posee y, sin embargo, estar equivocado al pretender usarlo como un genuino concepto de género natural atómico. Si esto es así, entonces este hablante no podría establecer *a priori* la verdad de la segunda premisa del argumento de Boghossian y tampoco podría extraer *a priori* la conclusión de la *reductio*. Brown se propone, en consecuencia, solucionar este problema planteando una nueva versión del argumento de la consecuencia.

Lo primero que debe señalarse es que, de los tres autores examinados en esta sección, Brown es, con mucho, quien aporta más evidencias textuales para fundamentar su argumento. En este sentido, atiende especialmente a una serie de indicaciones presentes en los trabajos de Burge acerca de las condiciones empíricas necesarias para que un agente pueda pensar contenidos individuados en función de factores externos. Su pretensión es averiguar qué otras situaciones podrían darse, aparte de la propia

presencia de H₂O en su entorno natural, para que un sujeto pudiera poseer contenidos mentales acerca del agua. De acuerdo con Burge, Óscar₂, el habitante de la Tierra Gemela, también podría poseer el concepto de agua o bien si tuviera los conocimientos científicos requeridos para teorizar sobre esa sustancia, o bien si careciera de estos conocimientos (es decir, si fuera agnóstico sobre las condiciones de aplicación del concepto) pero habitara en una comunidad lingüística que asocia determinados usos con el término “agua” (Burge 1982, pp. 96 y ss.). La lección que Brown extrae de estas indicaciones de Burge queda recogida en la siguiente implicación:

IMPLICACIÓN Q: Necesariamente, si un hablante tiene un pensamiento que involucra el concepto de un género natural K y es agnóstico sobre las condiciones de aplicación de dicho concepto, entonces (i) o bien este hablante está en un entorno que contiene K (ii) o bien este hablante es parte de una comunidad lingüística que posee el concepto de K.

La IMPLICACIÓN Q establece que un agente no experto en el dominio de un término de género natural “K” puede poseer el concepto de K o bien en caso de que haya entrado en contacto con ejemplares de K, o bien si ha aprendido el uso de ese concepto de otros hablantes. En cualquier caso, es digno de señalarse que, según la IMPLICACIÓN Q, la posesión de ese concepto requiere la existencia de entidades externas al hablante, bien se trate de la presencia física de K en el entorno o de la existencia de otros usuarios de su lenguaje. La necesidad de contar con al menos una de estas condiciones reside en el hecho de que sólo ellas nos permitirían determinar que el concepto que posee el hablante en cuestión es el concepto de género natural K y no cualquier otro concepto K* que le resulte indistinguible a este sujeto. Es decir, si un hablante no es capaz de diferenciar entre las condiciones de aplicación de dos conceptos K y K* (como Bert con respecto a la artritis y la artritis*), las pautas del externismo nos invitan a atribuir a este hablante sólo uno de ellos, K, bien en función de la presencia de K (y no K*) en su entorno, o bien a partir del hecho de que los usos lingüísticos de su comunidad fijan K (y no K*) como el concepto expresado. Sin embargo, Brown mantiene que aunque esta IMPLICACIÓN Q puede establecerse *a priori*, un hablante no experto es incapaz de deducir puramente *a priori* cualquiera de los consecuentes, (i) o (ii), a partir de su posesión del concepto K. La razón es que dicho hablante no puede averiguar *a priori* si ese concepto es un genuino concepto de género natural.²⁰³

²⁰³ Como ya indicó Boghossian, para determinar que el término de género natural “K” expresa un genuino concepto de género natural en los parámetros que requiere el partidario del externismo, es preciso

Análogamente, si tomáramos en consideración un concepto de género no natural N,²⁰⁴ nos encontraríamos con el mismo problema. De este modo, aunque es necesario que un sujeto pertenezca a una comunidad que tiene el concepto N en caso de que él posea ese concepto y sea agnóstico sobre sus condiciones de aplicación, dicho sujeto no puede saber *a priori* si N es un concepto de género no natural. Por este motivo, Brown refina lo establecido en IMPLICACIÓN Q ofreciendo una implicación distinta:

IMPLICACIÓN S: Necesariamente, si un hablante tiene un pensamiento que involucra el concepto C y este hablante es agnóstico sobre las condiciones de aplicación de C, entonces o este hablante (i) vive en un entorno que contiene instancias de C y C es un concepto de género natural o (ii) es parte de una comunidad que posee el concepto C, tanto si C es un concepto de género natural como si no lo es.

En efecto, si la dependencia que el externismo establece entre la posesión de un concepto y las condiciones empíricas relevantes pretende tener un carácter *a priori*, entonces no puede presuponerse que el sujeto sea capaz de saber de esa manera que el concepto relevante es de género natural o de género no natural. Lo único que se requiere para ello es que este hablante sepa que tiene un concepto C y que sea agnóstico sobre sus condiciones de aplicación. Como, según Brown, cualquier sujeto puede saber ambas cosas *a priori*, entonces cualquiera podría deducir, asimismo *a priori*, que o bien su entorno contiene instancias de C – en caso de que C sea un concepto de género natural –, o bien que existen otros hablantes de quienes ha aprendido el uso del término que expresa C.

A partir de la IMPLICACIÓN S podemos enunciar la *reductio* de Brown. La primera premisa, que puede establecer *a priori* cualquier agente que posea acceso privilegiado a sus propios conceptos, es la siguiente:

(RBr 1) Tengo el concepto C y soy agnóstico sobre sus condiciones de aplicación.

Podemos ofrecer, asimismo, una formulación en primera persona de la IMPLICACIÓN S, establecida también puramente *a priori*:

descartar que ese concepto no se aplica a una mezcla heterogénea o, simplemente, que ese concepto no carece de extensión. Y ninguna de estas cosas puede establecerse por medios puramente apriorísticos.

²⁰⁴ Brown entiende por “concepto de género no natural” un concepto expresado por un término general que no es un término de género natural.

(RBr 2) Si tengo el concepto C y soy agnóstico sobre sus condiciones de aplicación, entonces E [o bien mi entorno contiene ejemplares de C y C es un concepto de género natural o bien formo parte de una comunidad que posee el concepto C, tanto si C es un concepto de género natural como si no lo es].

Como vemos “E” abrevia la disyunción de las dos condiciones empíricas establecidas en IMPLICACIÓN S como necesarias para la posesión de C por parte de un hablante no-experto. Ahora bien, en virtud del conocimiento *a priori* de (RBr 1) y (RBr 2) podemos establecer la tesis E igualmente *a priori*. Pero, como dice Brown, resulta altamente contraintuitivo aceptar que seamos capaces de acceder *a priori* a cualquiera de las condiciones empíricas recogidas en E. En consecuencia, sólo queda rechazar el compatibilismo, dado que esta posición es la conjunción de las dos tesis (acceso privilegiado y externismo) que nos permiten deducir *a priori* la conclusión de la *reductio*.

Con la exposición de la *reductio* de Brown concluimos la reconstrucción de las tres versiones del argumento de la consecuencia, no sin antes indicar que los puntos más controvertidos del trabajo de esta autora son, en primer lugar, la idea de que el externismo es una tesis establecida puramente *a priori* y, en segundo lugar, ciertas tensiones internas al ejemplo que usa para ilustrar su *reductio*, tensiones que tendremos oportunidad de detallar en la siguiente sección.

2. Respuestas compatibilistas a las distintas versiones del argumento de la consecuencia

En lo que resta de capítulo pretendemos ofrecer un examen de las distintas objeciones que los autores compatibilistas han presentado contra el argumento de la consecuencia en cualquiera de sus variantes. Hemos optado por clasificar estas repuestas en tres grandes grupos. En primer lugar, hay autores que cuestionan la verdad y/o el carácter *a priori* de al menos una de las premisas del argumento. En segundo lugar, otros autores aceptan la verdad y aprioricidad de las premisas del argumento pero, al mismo tiempo, defienden que dicha consecuencia no tiene nada de implausible, sino que, por el contrario, proporcionaría una sugerencia novedosa e interesante para la construcción de argumentos anti-escépticos. Finalmente, hay autores que, por un lado, no cuestionan ni la verdad ni la aprioricidad de las premisas del argumento de la consecuencia pero que, por otro lado, consideran muy dudoso el potencial anti-escéptico del compatibilismo.

2.1 *Cuestionamiento de las premisas del argumento de la consecuencia*

Esta subsección está dedicada al análisis de aquellas alegaciones compatibilistas que cuestionan la verdad y/o el carácter *a priori* de al menos una de las premisas del argumento de la consecuencia, en alguna de sus tres versiones. Supuesta la validez formal del argumento, si se demostrara que alguna de sus premisas no es verdadera, se podría evitar la obtención de la conclusión incompatibilista. Asimismo, si se aceptara la verdad de las premisas pero no su conocimiento puramente *a priori*, entonces tampoco podríamos deducir *a priori* la conclusión, que es precisamente la consecuencia indeseable que se le imputa al compatibilismo.

2.1.1 *Contra el carácter a priori de las premisas del argumento de la consecuencia*

Comenzaremos examinando las objeciones de algunos autores, centradas en la noción de aprioricidad que manejan los defensores del argumento de la consecuencia en cualquiera de sus tres versiones. Como ya dijimos, constituye un presupuesto fundamental de esta clase de razonamientos que el acceso privilegiado a los propios pensamientos puede describirse como conocimiento *a priori* de los mismos. Sin embargo, es digno de mencionarse que tanto en el planteamiento simple como en los casos de traslado no se requería esta asunción y, de hecho, pudimos exponer y discutir a fondo estos argumentos sin comprometernos con una comprensión del acceso epistémico de primera persona en tanto que acceso *a priori* a los mismos.²⁰⁵

Por otra parte, existe un problema añadido con esta equivalencia entre la aprioricidad y el acceso privilegiado que se propone ahora desde el frente incompatibilista, a saber, que los defensores del argumento de la consecuencia mantienen una concepción muy restrictiva del conocimiento *a priori*. Debido a esto, muchas réplicas compatibilistas intentan mostrar, en primer lugar, que el acceso privilegiado a los contenidos mentales no posee los rasgos que estos autores incompatibilistas atribuyen al conocimiento *a priori* y, en segundo lugar, que ni siquiera las tesis externistas pueden establecerse *a priori* en el sentido requerido por los defensores del argumento de la consecuencia.

²⁰⁵ Recordemos, además, que el propio Burge en (1988a, p. 662) ya rehusaba describir el privilegio epistémico de primera persona como la capacidad de acceder *a priori* a los propios contenidos mentales.

Con respecto al primer punto de la réplica compatibilista, conviene comenzar recordando que McKinsey sostenía que el acceso privilegiado a los propios pensamientos tiene lugar *a priori*, en tanto que puede obtenerse (i) sin ninguna investigación empírica y (ii) sin ninguna presuposición sobre el mundo externo (McKinsey 1991, p. 9). Boghossian, por su parte, mantiene una concepción más débil y sólo exige el primer rasgo de la noción de *a priori* de McKinsey para mantener que el acceso privilegiado a los propios pensamientos tiene lugar *a priori* (Boghossian 1998, p. 198). Por último, Brown reconoce que la comprensión de la aprioricidad que interviene en los textos de McKinsey no resulta completamente afortunada pero, a pesar de ello, declara que ésa será la comprensión asumida en su propio argumento (Brown 1995, p. 185).

Sin embargo, desde el frente compatibilista se aduce que existe un serio problema con esta descripción de la aprioricidad presente en los trabajos de estos tres autores. Según señala Vilanova en (2009, p. 236) esta noción es tan exigente que resulta dudoso que alguno de nuestros conocimientos (incluido el conocimiento de los propios estados mentales) pueda ser obtenido de esa manera²⁰⁶ ya que, de acuerdo con dicha noción, una persona que se haya visto privada de percepciones sensoriales desde el comienzo de su vida podría saber *a priori* que cree que el agua es húmeda, puesto que para obtener ese conocimiento no se requieren investigaciones empíricas de ningún tipo. No obstante, este resultado es absurdo y merma la plausibilidad del argumento de la consecuencia. Por otro lado, Moya indica en (1998, p. 245) que en la *reductio* de Boghossian la concepción del acceso privilegiado verdaderamente operativa remite, no tanto al conocimiento *a priori* de los propios pensamientos como al conocimiento *a priori* del dominio de conceptos concretos. El problema es que si consideramos que los hablantes pueden decidir *a priori* que dominan un concepto determinado, estaríamos acercándonos peligrosamente a una concepción de la competencia conceptual *à la* Humpty Dumpty, en la cual se obviaría el papel que en este sentido deberían jugar los criterios socialmente establecidos para sancionar la aplicación de los conceptos y el uso de los términos.²⁰⁷

²⁰⁶ En este mismo sentido cf. Davies (2000), p. 408.

²⁰⁷ En (1998) Corbí insiste en la misma línea que Moya y mantiene que el dominio que un individuo tiene de un concepto determinado no es independiente de ciertos hechos públicos sobre él. Este autor compara el dominio de un concepto de género natural como el de agua con el dominio del concepto de hacer *jaque mate*. No resulta verosímil pensar que un hablante puede dominar este segundo concepto sin que haya tomado partido en ciertas prácticas (juegos, lecturas, conversaciones). Si se requieren esta clase de interacciones empíricas para dominar el concepto, entonces nadie puede decir *a priori* que lo domina. Lo

Dicho esto, es digno de señalarse que también se ha sometido a discusión un segundo punto en relación al uso de la noción de *a priori* presente en la argumentación incompatibilista. En este sentido, cabe cuestionar que la tesis general del externismo sea una tesis que se establece *a priori* en el sentido indicado o, incluso, que a pesar de serlo, pueda ser conocida *a priori* en el mismo sentido que obtenemos conocimiento *a priori* de nuestros conceptos, contenidos y pensamientos. Un cuestionamiento del primer tipo lo encontramos en Baker (2007). Esta autora defiende que la clase de justificaciones que sustentan las tesis del externismo (en especial las tesis del externismo social) tienen la forma de condicionales contrafácticos contingentes y que, por esta razón, no pueden asimilarse a intuiciones puramente *a priori* (Baker 2007, p. 295). Por su parte, en (1999) Nuccetelli adopta la segunda estrategia y sostiene que, aunque admitiéramos que el privilegio que caracteriza el acceso a nuestros pensamientos y conceptos puede ser entendido en términos de aprioricidad, podríamos preguntar si las tesis externistas quedan establecidas *a priori en el mismo sentido*. De este modo, si fuera cierto que conocemos *a priori* la verdad de las dos premisas del argumento de la consecuencia, pero no conociéramos ambas en el mismo sentido de *a priori*, entonces los defensores del argumento de la consecuencia estarían cometiendo una falacia de equivocidad. La argumentación de esta autora para defender su tesis discurre del siguiente modo. Según se ha indicado, la concepción de aprioricidad que manejan los autores incompatibilistas supone que un conocimiento es obtenido *a priori* si, al menos, no ha sido necesario contar con investigaciones empíricas para ello. Se presupone, pues, que ese rasgo es común al conocimiento introspectivo y al conocimiento obtenido a partir de los experimentos mentales externistas. Con todo, Nuccetelli argumenta que este “*a priori* mínimo” no consigue ocultar las diferencias que existen entre el conocimiento obtenido por una y otra vía. En particular, sostiene que el carácter no estrictamente empírico de la justificación que obtiene el externismo a partir de experimentos mentales no significa que no existan creencias empíricas de fondo operando en el diseño de estos experimentos (creencias relativas a la química, a la identidad molecular, a los hábitos ordinarios de atribución de creencias, etc.). Además, Nuccetelli defiende que la lectura externista de los escenarios contrafácticos de los experimentos no tiene un carácter puramente deductivo, sino que dicha tesis es inferida como la mejor explicación posible

importante, según Corbí, es que no existen demasiadas diferencias entre el concepto de hacer jaque mate y el concepto de agua y que, por tanto, tampoco es plausible creer que un individuo puede establecer *a priori* que posee este último concepto en el sentido requerido por Boghossian.

de los hechos expuestos en el escenario. De esta manera, la conclusión de la *reductio*, en realidad, no tendría nada de sorprendente. Sólo supondría extraer una conclusión empírica del acceso privilegiado y de una tesis, el externismo, cuyo establecimiento requiere contar con un buen número de asunciones empíricas.

En nuestra opinión, las observaciones de Baker y Nuccetelli gozan de bastante plausibilidad. La tesis externistas, en cualquiera de sus versiones, no son tesis que un escéptico sobre la existencia del mundo externo mantendría y ni siquiera parece posible ofrecer una reformulación aceptable de las mismas que no se comprometa con la existencia de un mundo extra-mental.²⁰⁸ Dicho de otra manera, al defender que es preciso atender a factores externos, de carácter natural o social, para individuar algunos de nuestros pensamientos, contenidos y conceptos, estamos contando ya con la existencia de estos componentes mentales y con los aspectos del mundo externo que los determinan. A lo largo de este capítulo tendremos ocasión de volver sobre este punto.

2.1.2 *Cuestionamiento de la verdad de las premisas del argumento de la consecuencia*

Una vez expuestas las objeciones compatibilistas a la noción de aprioricidad presupuesta en el argumento de la consecuencia, resulta conveniente atender a una segunda estrategia para combatir dicho argumento, a saber, demostrar que la comprensión del externismo que figura en él es inapropiada. En concreto, lo que se niega es que las tesis externistas, entendidas de una manera cabal, den lugar a instancias apriorísticas como las que figuran en la segunda premisa de las distintas versiones del argumento de la consecuencia. Recordemos cuáles eran dichas premisas en cada caso.

En la paradoja de McKinsey:

(McK 2) La proposición de que Óscar₁ está pensando que el agua es húmeda implica conceptualmente E [hay H₂O en su entorno].

En la *reductio* de Boghossian:

(RB 2) Si tengo el concepto de agua [y es un concepto atómico que pretende nombrar un concepto de género natural], entonces el agua existe.

²⁰⁸ Pensemos en una reformulación de esta clase como, por ejemplo, “en caso de que exista un mundo externo que contiene géneros naturales y hablantes que asignan convenciones semánticas a los términos que uso, la identidad de mis pensamientos dependería de ellos”. En realidad nadie ha comprendido el externismo desde ese punto de vista, sino que en las formulaciones usuales de esta tesis se parte de la existencia de aquellos factores externos que correlacionamos con los pensamientos propios y *ajenos*. Cf. Beebe (2001, pp. 367 y ss).

En la *reductio* de Brown:

(RBr 2) Si tengo el concepto C y soy agnóstico sobre sus condiciones de aplicación, entonces E [o bien mi entorno contiene ejemplares de C y C es un concepto de género natural o bien formo parte de una comunidad que posee el concepto C, tanto si C es un concepto de género natural como si no lo es].

2.1.2.1 *La verdad de las premisas de la paradoja de McKinsey*

En lo que se refiere al argumento de McKinsey, uno de los defectos más notorios de su trabajo es su falta de atención a aquellos textos externistas que, como Burge (1982), advierten expresamente que “E” no sólo puede traducirse por “hay H₂O en el entorno de Óscar₁”, sino que un agente podría mantener pensamientos sobre el agua sin que existiera agua en su entorno, siempre y cuando hubiera adquirido el concepto por vías indirectas, tales como su capacidad de teorizar sobre ese líquido o su inserción en una comunidad con usos lingüísticos asociados al término “agua”.²⁰⁹ Sin embargo, este error tiene una importancia relativa porque, tal y como ha mostrado Brown, es posible modificar el argumento para incluir en sus premisas estas condiciones de adquisición conceptual.

Igualmente puede cuestionarse un segundo aspecto de la comprensión del externismo que exhibe este autor y es que McKinsey describe como una relación de “implicación conceptual” la clase de relación que el externismo postula entre un pensamiento y E, el factor externo que determina su contenido. La clave de este tipo de relaciones, tal y como McKinsey las describe, es que permiten deducir *a priori* el hecho empírico E a partir de la posesión (congnoscible *a priori*) de un pensamiento. Sin embargo, tanto Brueckner en (1992), (2004) y (2007) como Goldberg en (2003b) han discutido este modo de entender el externismo y han defendido que la clase de dependencia que esta tesis reconoce entre el pensamiento de Óscar₁ y E es una dependencia de naturaleza metafísica (no necesariamente cognoscible *a priori*) y no una dependencia conceptual o lógica como McKinsey pretende. Así, aunque el externismo mantiene que la proposición de que Óscar₁ está pensando que el agua es húmeda requiere que haya H₂O en su entorno, puesto que la identificación “el agua es H₂O” que subyace al lema externista de (McK 2) es sólo cognoscible *a posteriori*, entonces la

²⁰⁹ Cf. Brueckner (2004), p. 378 y (2007), p. 40.

clase de dependencia que el externismo establece es, asimismo, una dependencia que sólo podemos descubrir *a posteriori*.²¹⁰

Antes de dar por válidas estas réplicas a la paradoja de McKinsey, vamos a retomar el razonamiento que este autor presentó en (1991) para defender que la dependencia externista entre nuestros pensamientos y los factores empíricos que condicionan sus contenidos no puede ser descrita como una dependencia de carácter metafísico, sino conceptual o lógico. En lo que sigue, ofreceremos un argumento para cuestionar dicho razonamiento y hacer plausible la tesis de que es perfectamente posible entender el externismo como una tesis metafísica sin trivializarla. Para ello proponemos, en primer lugar, reconstruir el razonamiento de McKinsey de la siguiente manera:

- (1) Supongamos que el externismo mantiene que la dependencia que existe entre nuestros pensamientos y los factores externos que determinan sus contenidos es de carácter metafísico.
- (2) Si existe una dependencia metafísica entre nuestros pensamientos y los condicionantes de sus contenidos, entonces nuestros pensamientos presuponen metafísicamente estos condicionantes externos.
- (3) Los condicionantes externos son individuos distintos de aquél al que se le atribuye un estado psicológico.
- (4) Un pensamiento es un estado psicológico en sentido amplio si su atribución a un hablante *presupone la existencia* de un individuo distinto de aquél al que se le atribuye el pensamiento.

²¹⁰ Por otro lado, en (2004) y (2007) Brueckner defiende que este último razonamiento también serviría para combatir una versión más sofisticada de la propuesta de McKinsey – cercana a la argumentación de Brown – que recogiera también entre las condiciones de posesión del pensamiento de que el agua es húmeda, la capacidad de teorizar sobre el H₂O y la presencia de hablantes en el entorno que asocian con el término “agua” ciertas convenciones. La premisa modificada rezaría así:

(McK 2*) La proposición de que Óscar₁ está pensando que el agua es húmeda requiere (i) que exista H₂O en su entorno o (ii) que Óscar₁ sea capaz de teorizar sobre el H₂O o (iii) que Óscar₁ habite en una comunidad lingüística algunos de cuyos miembros teorizan sobre el H₂O.

Dado que, de nuevo, la identificación “el agua es H₂O” está presupuesta en la enunciación de esta premisa y dicha identificación expresa una verdad necesaria *a posteriori*, entonces la verdad de (McK 2*) no es accesible *a priori*, al contrario de lo que sostienen McKinsey y Brown. En (1998a, pp. 367 y ss.) McLaughlin y Tye ofrecen una réplica al argumento de McKinsey bastante similar. De acuerdo con estos autores, las tesis del externismo natural establecen que se requiere contar con factores externos para individuar nuestros conceptos de género natural. En concreto, se apoyan en la variante del externismo natural que respalda McGinn en (1989), quien afirma que esos factores externos son aquellos con los que el sujeto interactúa causalmente. Ahora bien, de ser así, entonces sólo podemos deducir *a priori* que existe el factor empírico que determina la identidad de uno de nuestros conceptos si podemos saber *a priori* que poseemos este concepto y si, además, podemos saber *a priori* que ese concepto es un genuino concepto de género natural. Sin embargo, esto último no es cognoscible *a priori* y, en consecuencia, tampoco es cognoscible *a priori* la existencia del determinante empírico relevante.

- (5) Por (1)-(4), el externismo defiende que nuestros pensamientos son estados psicológicos en sentido amplio.
- (6) Todo hablante presupone metafísicamente la existencia de sus progenitores.
- (7) Todo pensamiento que se le atribuye a un hablante presupone metafísicamente la existencia de los progenitores del hablante, considerados como individuos distintos de aquél al que se le atribuye el estado.
- (8) Todos los pensamientos están individuados de manera externista.

Como vimos, deducir (8), la aplicabilidad general del externismo a todos nuestros pensamientos, no supone para McKinsey establecer una conclusión favorable al externismo, sino reducir esta tesis a una mera trivialidad.²¹¹

No obstante, pensamos que algunos aspectos del razonamiento de McKinsey son sumamente cuestionables. En nuestra opinión, la clave para trivializar el externismo entendido como una tesis metafísica (y no lógica o conceptual) es el uso que hace este autor de la noción de estado psicológico en sentido amplio de Putnam y la caracterización del externismo que figura en la premisa (5) y de los factores externos que figura en (3). A este respecto, McKinsey debe afrontar una primera dificultad y es que el uso de dicha noción no es demasiado frecuente entre los partidarios del externismo natural o social y creemos que existen al menos dos buenas razones para ello. La primera, recogida por Burge en (2007, p. 13), es que los autores externistas son, en cuanto tales, defensores de una concepción unificada del contenido, en el sentido que Bilgrami da a esta noción en (1994). De este modo, como los externistas rechazan bifurcar el contenido de nuestros estados mentales (y los estados mentales mismos) en contenido (y pensamientos) en sentido amplio y en sentido estrecho, no tiene demasiada utilidad caracterizar los pensamientos individuados de manera externista como estados mentales con contenido en sentido amplio, al no poder establecer el contraste correspondiente. La segunda razón es que la noción de Putnam de estado psicológico en sentido amplio resulta excesivamente vaga para los propósitos del externista. Recordemos que McKinsey usa esta noción con el fin de caracterizar los condicionantes externos de los contenidos de nuestros estados psicológicos (en sentido amplio) en tanto que se trata de (i) otros individuos distintos del sujeto de atribución, según la premisa (3), y de (ii) individuos cuya existencia es necesaria para la atribución del pensamiento

²¹¹ Es decir, que este autor intenta demostrar lo absurdo de concebir el externismo como una tesis metafísica argumentando que, de ese modo, dicha tesis conllevaría que todos los pensamientos estarían indiscriminadamente individuados de manera externista, al depender cualquier pensamiento que se atribuya a un individuo de la existencia de los padres de éste.

a un sujeto, atendiendo a la premisa (4). Los progenitores del sujeto de una atribución psicológica serían condicionantes externos según esta caracterización, puesto que cumplen los requisitos (i) y (ii). Sin embargo, esto no dice nada en contra del externismo, sino en contra de la caracterización del mismo que ofrece McKinsey. En lo que sigue trataremos de justificar esta afirmación arguyendo que dicha caracterización incluiría como condicionantes externos de los contenidos de nuestros pensamientos a elementos que no intervienen en absoluto en la individuación de los mismos de acuerdo con una comprensión razonable de las tesis externistas.

El núcleo de nuestro argumento viene dado por estas dos premisas. En primer lugar, el externismo psicológico aboga por la dependencia de los contenidos de los pensamientos de un hablante con respecto a factores externos *cuya variación contrafáctica conllevaría una variación en la identidad de sus contenidos*. En segundo lugar, los progenitores de un hablante no son condicionantes externos de esa clase, ya que si estipuláramos una variación contrafáctica de modo tal que los progenitores de un sujeto de atribución (considerados como factores externos) fueran individuos distintos, lo único que obtendríamos sería una variación en la identidad de dicho sujeto y, sin embargo, no se habría dicho nada aún sobre la identidad del contenido de sus pensamientos.

Como punto de partida, proponemos tomar de nuevo en consideración los experimentos mentales externistas para comprobar que una variación de esa clase, que afectara únicamente a la identidad del sujeto de atribución, sería absolutamente neutral en lo que concierne a la cuestión de la identidad de los contenidos mentales de un agente. Pensemos, en primer lugar, en el experimento original de la Tierra Gemela propuesto por Putnam. Una de sus premisas es que Óscar₁ y Óscar₂ son *hablantes diferentes*, aunque exactamente idénticos en todos sus aspectos internos y que, en tanto que hablantes distintos, nacieron de diferentes progenitores. En segundo lugar, el experimento de Burge en torno a la noción de artritis estipula una variación en el entorno externo de *un único sujeto*, Bert, al que atribuimos un determinado pensamiento para preguntar, a continuación, si ese pensamiento tendría el mismo contenido o no en una hipotética situación contrafáctica.²¹² Según vemos, resulta indiferente que en los

²¹² El hecho de que un externista opte por la segunda opción y diseñe un test que tome en consideración al mismo individuo en una situación actual y en una situación alternativa es sólo un recurso para maximizar el control de las variables puestas en juego en el experimento. En este sentido, si examinamos la situación de un mismo individuo en dos escenarios resulta más sencillo postular la identidad de todos sus factores

experimentos mentales que se emplean usualmente a favor del externismo se hable de dos hablantes distintos (aunque idénticos en sus aspectos internos) o de un mismo hablante en una situación actual y en una situación alternativa, lo cual apoya la idea de que una variación contrafáctica que afecte a la identidad de los progenitores de un agente y que suponga un cambio en la identidad de éste sería absolutamente irrelevante para los propósitos del externismo, puesto que aún quedaría pendiente la pregunta de si ese hablante contrafáctico (distinto al de la situación actual) amerita o no la atribución de los mismos contenidos mentales que el hablante de la situación actual.

La razón fundamental para apreciar la irrelevancia de esta clase de factores es la siguiente. Tanto el externismo como el internismo están interesados en detectar cuáles son las condiciones de identidad de los contenidos de nuestros pensamientos. Es decir, ambos enfoques pretenden determinar qué hace que un pensamiento tenga la identidad que tiene (que sea el pensamiento de que P y no de que Q o R) en cuanto que su identidad viene determinada, en parte, por la identidad de su contenido.²¹³ Por esta razón, la clave de los experimentos mentales a favor del externismo es la premisa que postula una identidad intrínseca, bien entre los diferentes sujetos que los protagonizan, bien en el mismo sujeto en ambas situaciones (actual y contrafáctica). Si a pesar de esa identidad intrínseca, nuestras intuiciones nos invitan a atribuir contenidos psicológicos (y pensamientos) diferentes en cada situación, habremos efectuado una lectura del experimento favorable al externismo.

La conclusión que cabe extraer de lo anterior es que entender el externismo como una tesis sobre la dependencia metafísica (no conceptual o lógica) de nuestros pensamientos es bastante distinto de defender, como según McKinsey hace el externismo, que los pensamientos que atribuimos a un hablante dependen de elementos externos cuya existencia es necesaria para la atribución de pensamientos a *ese sujeto*. Lo que se requiere, más bien, es detectar qué factores determinan la atribución de *esos* pensamientos (pensamientos con tal o cual contenido) a *cualquier* sujeto. Por esta razón, si cabe entender el externismo como una tesis metafísica, al contrario de lo que este

internos y confiar en que no se ha producido una variación subrepticia en algunos de estos factores que explique la diferencia entre estados mentales que el externista quiere concluir.

²¹³ Tal y como señalamos en la introducción, la identidad de un pensamiento depende de dos factores, a saber, (i) el tipo de actitud psicológica del sujeto hacia el contenido del pensamiento y (ii) la identidad del propio contenido. La discusión entre partidarios del internismo y del externismo en psicología acerca de las condiciones de identidad de nuestros pensamientos versa fundamentalmente acerca de la naturaleza interna o, al menos, parcialmente externa de los determinantes del contenido proposicional de dichos contenidos, más bien que acerca de los condicionantes del otro componente de un pensamiento (el tipo de actitud psicológica).

autor mantiene, entonces no está tan claro que la segunda premisa de la paradoja de McKinsey sea, en realidad, una instancia de las tesis externistas.

2.1.2.2 *La verdad de las premisas de la reductio de Boghossian*

Una vez que hemos tomado en consideración estas objeciones al establecimiento *a priori* de la segunda premisa, (McK 2), del argumento de McKinsey, nos ocuparemos de la premisa que Boghossian propone en su lugar, (RB 2):

(RB 2) Si tengo el concepto de agua [y es un concepto *atómico* que *pretende* aplicarse a un género natural], entonces el agua existe.

La aclaración que incluimos entre corchetes, como se vio, es de extrema importancia. Según argumentaba Boghossian, para demostrar que el externismo implica que aquellos pensamientos que expresamos haciendo uso de oraciones que incluyen un término de género natural son pensamientos dependientes de objeto, basta con que el concepto expresado por dicho término sea (i) un concepto *atómico* y (ii) un concepto que *pretende* aplicarse a un género natural. Especificar este segundo requisito es una forma de evitar una objeción inmediata a su *reductio*, a saber, que nadie puede saber *a priori* que uno de sus conceptos es un concepto que se aplica de hecho a un género natural existente. En caso de que Boghossian exigiera que el concepto de agua que figura en el antecedente de (RB 2) fuera un concepto genuino de género natural,²¹⁴ dicha premisa no podría establecerse *a priori*, ni, consiguientemente, podría establecerse de ese modo la conclusión del argumento, por lo que la *reductio* perdería su efectividad.

Por su parte, el requisito de la atomicidad tiene un papel esencial en la derivación de la dependencia de objeto a partir del externismo. Para explicar su importancia, vamos a enumerar los rasgos de la dicotomía atomicidad/composicionalidad que Boghossian presenta en su trabajo:

(A/C 1) Un concepto compuesto está definido por otros conceptos, mientras que un concepto atómico no lo está (Boghossian 1998, p. 201).

(A/C 2) La atomicidad o composicionalidad de un concepto es función exclusiva de su sintaxis interna. Un concepto será compuesto si está definido por otros conceptos. Si no, será atómico (Boghossian 1998, p. 208).

²¹⁴ Entendemos por “concepto genuino de género natural” lo mismo que Boghossian, a saber, un concepto que no carece de extensión y cuya extensión no es una mezcla heterogénea.

(A/C 3) De mantener el acceso *a priori* a los propios conceptos, un hablante es capaz de saber *a priori* si uno de sus conceptos es atómico o compuesto (Boghossian 1998, p. 210).

(A/C 4) El externista requiere presuponer que el concepto que Óscar₁ y Óscar₂ expresan haciendo uso del término “agua” es un concepto atómico (Boghossian 1998, p. 201).²¹⁵

En lo siguiente vamos a examinar qué respuestas han ofrecido diversos autores compatibilistas para evitar el establecimiento *a priori* del antecedente de la premisa (RB 2) en relación a la atomicidad del concepto de agua. En nuestra opinión, la posición de Corbí en (1998) es una de las más interesantes. De acuerdo con este autor, la cuestión crucial para admitir (o no) la verdad de esta premisa es la respuesta que se dé a la siguiente pregunta: ¿podemos saber *a priori* cuáles de nuestros conceptos están individuados en función de factores externos? Mientras que el externista responde de manera negativa a esta cuestión, Boghossian insiste en que, en la medida en que podamos saber *a priori* cuáles de nuestros conceptos son atómicos y pretenden aplicarse a un género natural, entonces podremos saber *a priori* que tales conceptos recibirán una individuación externista. Sin embargo, Corbí señala que la respuesta de Boghossian esconde una cierta petición de principio y la razón es que para este autor un concepto atómico, en realidad, no es más que un concepto individuado en función de factores externos.

El argumento que Corbí emplea para evidenciar la existencia de esta suposición en el trabajo de Boghossian es el siguiente. De acuerdo con la justificación que se ofrece para el rasgo (A/C 4) del par atómico/compuesto, Boghossian afirma que la lectura externista del escenario de la Tierra Gemela requiere descartar que los conceptos que expresan Óscar₁ y Óscar₂ al usar el término “agua” sean conceptos compuestos. La razón, como vimos, es que los conceptos compuestos, en cuanto tales, están determinados por otros conceptos internos al hablante, en función de los cuales se

²¹⁵ Para comprobarlo, asumamos provisionalmente la tesis opuesta, esto es, que los conceptos relevantes de los gemelos son compuestos, para entender que las conclusiones extraídas serían contrarias a las pretensiones externistas. Así, de tratarse de conceptos composicionales, dichos conceptos estarían compuestos por otros más simples como los conceptos de líquido, inodoro, incoloro e insípido. Pero, en la medida en que forma parte de las premisas del experimento mental que ese repertorio de conceptos es poseído por ambos individuos (duplicados internos idénticos), Óscar₁ y Óscar₂ estarían expresando el mismo concepto al usar el término “agua”, contrariamente a lo que pretende el externista. De esta manera, Boghossian afirma que “el experimento presupone que el agua no puede ser pensada como susceptible de ser definida como: un líquido insípido, inodoro que fluye en los ríos y lagos. Si fuera un concepto composicional de esta clase, su extensión quedaría determinada por la extensión de sus partes ingredientes” (Boghossian 1998, p. 201).

definen. Como Óscar₁ y Óscar₂ son, por hipótesis, hablantes intrínsecamente idénticos, entonces estarán expresando el mismo concepto mediante el uso del término “agua” en caso de que lo concibiéramos como un concepto compuesto, dado que uno y otro poseen en su repertorio conceptual (todos) los conceptos básicos que componen el supuesto concepto compuesto de agua. Ahora bien, el problema es que, en último término, el externista mantendrá que todo concepto que reciba una individuación externista (como aquellos que expresamos haciendo uso de términos de género natural) será un concepto atómico porque, por definición, esta clase de conceptos no están individuados exclusivamente en función de factores internos como son otros conceptos más básicos. Pero esto conlleva que hablar de conceptos atómicos es, en definitiva, hablar de conceptos individuados de manera externista.²¹⁶

En Moya (1998) encontramos otro argumento similar dirigido a impedir el establecimiento *a priori* del antecedente de la premisa (RB 2) que pasa por cuestionar la idea de Boghossian de que podemos saber *a priori* qué conceptos (que pretenden aplicarse a un género natural) son conceptos atómicos. Tal y como se vio en el capítulo anterior, Moya aboga por una concepción más refinada de la individuación externista que la que propone el externismo estrictamente causal, siendo ésta una versión del externismo cercana a la que emplea Boghossian en su *reductio*. De acuerdo con esta versión, los factores del entorno externo involucrados en la historia causal de un sujeto intervienen en la individuación de sus conceptos y contenidos mentales. Por su parte, la comprensión del externismo de Moya, que él denomina en (1998, p. 248) “externismo normativo”, concreta la intervención de los factores externos en la individuación de nuestros conceptos del siguiente modo, a saber, aquellas instancias que han figurado en calidad de paradigmas en el aprendizaje del uso de nuestros términos de género natural determinarán qué conceptos expresamos mediante su uso.

Como partidario del externismo normativo, Moya examina qué papel desempeña el par atomicidad/composicionalidad en su propia teoría. Según nos dice este autor, es probable que exista una *presuposición* de atomicidad en relación a los conceptos que expresamos haciendo uso de términos de género natural y que dicha presuposición sea

²¹⁶ Podemos enunciar de otra manera la objeción de Corbí: la caracterización del par atómico/compuesto que ofrece Boghossian contiene rasgos inconsistentes. Por un lado, de acuerdo con el rasgo (A/C 2), los hechos sobre la atomicidad o composicionalidad sobrevienen a las propiedades internas de un individuo. Sin embargo, esto entra en tensión con la justificación de Boghossian para el rasgo (A/C 4), dado que en ella se presume que sólo la identidad de los conceptos compuestos sobreviene a las propiedades intrínsecas de los individuos, al estar determinados por los conceptos que los definen, pero que esto no ocurre con los conceptos atómicos, los cuales están determinados, en parte, por factores externos a los individuos.

cognoscible *a priori*. Sin embargo, esto no significa que pueda saberse *a priori* que tal presuposición se cumple, esto es, que esos conceptos son, de hecho, conceptos atómicos y no composicionales. El razonamiento de Moya para ofrecer esta distinción es el siguiente. De acuerdo con este autor, al usar términos de género natural pretendemos que los ejemplares que pertenecen a la extensión de esta clase de términos posean una “naturaleza” o “esencia unitaria” (Moya 1998, p. 249). Sin embargo, aunque esta pretensión existe, el conocimiento efectivo de esa esencia o naturaleza es tarea de la ciencia y, por tanto, se requiere una cantidad indeterminada de investigación empírica para lograr esta clase de conocimiento. Lo relevante es que Moya entiende la presuposición de atomicidad de un concepto únicamente como la pretensión de que dicho concepto se aplique a un género natural cuyas instancias posean una esencia unitaria, esto es, una misma propiedad estructural en el sentido de Putnam.

La objeción que este autor plantea a la tesis de Boghossian es la siguiente. Aceptando el externismo normativo, se presupone *a priori* que aquellos conceptos que se aplican a géneros naturales tienen, precisamente por pretender ser aplicados de este modo, un carácter atómico, pero averiguar si dicha presuposición tiene fundamento en la realidad no es algo que pueda dirimirse puramente *a priori*. En el caso de que nuestras indagaciones empíricas nos llevaran a concluir que el género natural en cuestión no está dotado de una esencia unitaria (si es vacío o si posee una composición heterogénea), la presuposición no se cumplirá y, por tanto, el concepto no será un genuino concepto de género natural – en el sentido de Boghossian – ni recibirá una individuación dependiente del entorno. Por estas razones, Moya afirma que la presuposición de atomicidad (no la atomicidad misma) es, en efecto, cognoscible *a priori*, pero que, dado que se trata de una mera presuposición, resulta falible. Si este autor está en lo correcto, entonces Boghossian no puede sostener que somos capaces de conocer *a priori* cuáles de nuestros conceptos que pretendemos aplicar a géneros naturales son conceptos atómicos.

Las consideraciones de Corbí y Moya ponen de manifiesto que la distinción entre conceptos atómicos y conceptos compuestos que figura en la *reductio* resulta un tanto problemática. Desde luego, sorprende que un autor como Boghossian, que analiza con tanto cuidado las premisas del experimento mental de la Tierra Gemela, haya optado por introducir una distinción que, como se ha visto, no figura en la formulación original de dicho experimento ni en las reconstrucciones del mismo que suelen ofrecer los partidarios del externismo psicológico. En principio, no sería reprochable introducir

esta dicotomía si con ello se clarificaran algunos de los supuestos de dicho experimento, pero hay razones para pensar que esta distinción presenta tensiones internas. Así, por un lado, este autor describe la atomicidad conceptual como un requisito para efectuar una lectura externista del escenario de la Tierra Gemela, por lo que la distinción entre conceptos atómicos y conceptos compuestos es una distinción que sólo se presentaría aceptando un marco externista de individuación. Sin embargo, por otro lado, Boghossian sostiene que la atomicidad o composicionalidad de nuestros conceptos depende de la sintaxis interna de éstos y, por tanto, se trata de un dato accesible *a priori* para cualquiera que mantenga que poseemos acceso *a priori* a nuestros conceptos. Pero al expresarse de esta manera, Boghossian supone que el hecho de que un concepto esté individuado de manera externista (o no) es algo que un agente debe poder decidir puramente *a priori* en caso de que tenga acceso privilegiado a sus conceptos, y esto es, justamente, lo que no puede decidirse *a priori*, según externistas como Corbí y Moya.

Una vez examinadas las réplicas compatibilistas que conciernen a la posibilidad de conocer *a priori* el carácter atómico de nuestros conceptos, vamos a tomar en consideración los problemas relativos al experimento mental de la Tierra Seca. A partir de dicho escenario, recordemos, Boghossian planteaba un dilema, a saber, o bien el concepto que expresan los habitantes de la Tierra Seca mediante el uso del término “agua” es un concepto compuesto, o bien es un concepto atómico. Este autor defiende que optar por la composicionalidad o por la atomicidad del presunto concepto nos llevaría, igualmente, a un camino sin salida. La conclusión que pretende extraer a partir de esto es que el externismo está comprometido con la tesis de que un agente que usa un término de género natural vacío no puede estar expresando ningún concepto. Si esto es así, y asumiendo, por mor del argumento, que somos capaces de determinar *a priori* si uno de los conceptos que pretendemos aplicar a un género natural es de carácter atómico, entonces podemos conocer *a priori*, en virtud de lo establecido a partir del escenario de la Tierra Seca, que ese concepto no carece de extensión. Pero esto nos llevaría directos a la conclusión de la *reductio*, puesto que podríamos conocer *a priori* una cuestión empírica, a saber, que existen aquellos factores externos que pertenecen a la extensión del género natural y que, de acuerdo con el externismo, determinan la identidad de nuestros conceptos.

En nuestra opinión, cabe emplear las conclusiones que se extrajeron del examen del par atómico-compuesto en Boghossian para debilitar el dilema que este autor propone con respecto a la situación de Óscar₃ en la Tierra Seca. Como vimos, el retrato

que ofrece Boghossian de la interpretación externista del escenario de la Tierra Gemela muestra que el carácter compuesto de un concepto conlleva una individuación exclusivamente interna del mismo – en función de sus conceptos componentes –o, dicho a la inversa, si los conceptos que Óscar₁ y Óscar₂ son atómicos es porque reciben una individuación en parte externa. De aceptar esto, un duplicado interno de estos sujetos como es Óscar₃ no tendría por qué expresar un concepto atómico al usar el término “agua” si no hay ningún factor externo que contribuya a su individuación. Por esta razón, este hablante bien podría expresar un concepto compuesto, cuyas condiciones de satisfacción serían más amplias que las condiciones de satisfacción de los conceptos que Óscar₁ y Óscar₂ expresan haciendo uso del término relevante. Al no existir ningún líquido en el entorno de la Tierra Seca que posea las propiedades que Óscar₃ atribuye a lo que denomina “agua”, no hay ningún factor que restrinja las condiciones de aplicación del término, bien al H₂O, bien al XYZ.²¹⁷

2.1.2.3 *La verdad de las premisas de la reductio de Brown*

En este último subapartado, nos ocuparemos de la segunda premisa del argumento de Brown, para lo cual comenzaremos tomando en consideración la estrategia que McLaughlin y Tye desarrollaron en (1998b) para impedir el establecimiento de dicha premisa. Recordemos primeramente su contenido:

²¹⁷ Algunos autores compatibilistas han ofrecido otro tipo de argumentos contra la lectura que realiza Boghossian del escenario de la Tierra Seca. Así, Moya rechaza en (1998, pp. 250 y ss.) una de las premisas que figuran en la refutación de la opción atómica, en particular, la premisa de que la oración “el agua es húmeda” no expresa ninguna proposición cuando es proferida por un habitante de la Tierra Seca. De acuerdo con Moya, el análisis de las condiciones de verdad de dicha oración cuando es proferida en ese entorno viene dado por el siguiente bicondicional: “El agua es húmeda” si y sólo si el agua es húmeda. Según dice Moya, un externista no tiene por qué sentirse obligado a afirmar que el miembro derecho de este bicondicional carece de condiciones de verdad. Es decir, podría mantener que es falso pero que tiene significado. De este modo, la oración “el agua es húmeda” tendría condiciones de verdad en la Tierra Seca aunque las peculiaridades de ese entorno impedirían que dichas condiciones fueran satisfechas. El concepto expresado mediante el término “agua” por los habitantes de la Tierra Seca sería, en palabras de este autor, “el concepto que los habitantes de la Tierra Seca tendrían si hubiera de hecho un género natural, con la apariencia externa apropiada, denotado allí por las preferencias de ‘agua’” (Moya 1998, p. 218). No obstante, creemos que la réplica de Moya resulta un tanto insatisfactoria. Decir que el concepto que expresan los habitantes de la Tierra seca al usar el término “agua” es el que expresarían si ese término tuviera extensión es admitir que, de hecho, no están expresando ningún concepto, por mucho que en una situación contrafáctica lo hicieran. Y admitir que un individuo no expresa ningún concepto cuando emplea un término de género natural carente de extensión es dar la razón a Boghossian en cuanto que supone admitir que los pensamientos individuados en función de factores externos son dependientes de objeto. Sin embargo, las consideraciones precedentes de Moya sobre la atomicidad conceptual le bastarían para esquivar la paradoja. De este modo, si no podemos saber *a priori* si un concepto atómico que pretende ser aplicado a un género natural tiene extensión, la conclusión de la *reductio* no puede ser derivada *a priori*, aunque se le conceda a Boghossian que esa clase de conceptos no pueden ser vacíos.

(RBr 2) Si tengo el concepto C y soy agnóstico sobre sus condiciones de aplicación, entonces E [o bien mi entorno contiene ejemplares de C y C es un concepto de género natural o bien formo parte de una comunidad que posee el concepto C, tanto si C es un concepto de género natural como si no lo es].

Según vimos, el razonamiento que ofrece Brown para establecer esta premisa (RBr 2) es que si mantenemos que un sujeto agnóstico sobre las condiciones de aplicación de un concepto es capaz de saber *a priori* que lo posee (aun cuando no sepa determinar *a priori* si es un concepto de género natural o no) podrá saber *a priori*, al menos, que pertenece a una comunidad de hablantes que establecen reglas para usar el concepto.

La réplica que han ofrecido McLaughlin y Tye a este argumento es que el ejemplo que Brown diseña para establecer la premisa (RBr 2) no demuestra, en realidad, que un hablante que posee un concepto de género no natural y es agnóstico sobre sus condiciones de aplicación debe estar inserto en una comunidad lingüística para poseer ese concepto. Vamos a examinar la réplica de estos autores citando el ejemplo de Brown y advirtiendo previamente de que los subrayados son nuestros:

“El argumento que ofrece Burge para la tesis de que [Adam, que es ignorante sobre la naturaleza del agua] no podría tener un pensamiento acerca del agua a menos que exista el agua o existan otros hablantes, parece que puede ser aplicado a pensamientos que involucran conceptos de género no natural. Imagine que [Adam] es agnóstico sobre las condiciones de aplicación del término ‘sofá’. Por ejemplo, puede aplicarlo de manera firme y correcta a lo que nosotros denominamos ‘sofás’, aunque no está seguro de si también se aplica a sillones amplios de un solo asiento. De acuerdo con Burge, si [Adam] es parte de una comunidad de hablantes del inglés entonces, a pesar de su agnosticismo, él tiene pensamientos que involucran el concepto de sofá. Pero si, contrafácticamente, [Adam] ha sido parte de una comunidad en la cual ‘sofá’ se aplica tanto a lo que llamamos ‘sofás’ como a sillones amplios de un solo asiento, entonces [Adam] podría haber tenido pensamientos sobre los ‘chofás’, donde el concepto de chofá se aplica tanto a lo que llamamos sofás como a esa clase de sillones. Ahora imagina que no hay otros hablantes en el entorno de Adam. ¿Cómo podría tener actitudes proposicionales que involucran el concepto de sofá? Puesto que ese concepto no es de género natural, el entorno de [Adam] no puede ayudarle a adquirir el concepto. No hay otros hablantes. Nada parece mostrar que esas

actitudes involucran el concepto de sofá en vez del concepto de chofá” (Brown 1995, p. 190).

A partir de este ejemplo Brown considera probado que un hablante agnóstico sobre las condiciones de aplicación de un determinado concepto de género no-natural, no podría poseer dicho concepto a menos que existiera una comunidad lingüística de referencia que ayudara a determinar su identidad, pero, según señalan McLaughlin y Tye, el ejemplo de Brown demuestra justamente lo contrario. El núcleo de su argumento descansa en la primera observación que hemos subrayado, a saber, que Adam aplica de manera firme y *correcta* el término “sofá”. Ahora bien, para estos autores, “si la palabra puede ser aplicada correctamente, entonces es que expresa un concepto” (McLaughlin/Tye 1998, p. 212). Es decir, lo que McLaughlin y Tye alegan es que forma parte del diseño del ejemplo que Adam está aplicando un término *de acuerdo con ciertas normas*, en la medida en que se juzgan estas aplicaciones como correctas. Pero, si esto es así, entonces un partidario del externismo social no necesitaría ningún requisito adicional para establecer que este sujeto está expresando un concepto. De acuerdo con Burge, “los conceptos marcan habilidades”, en concreto, la habilidad de usar un término con relativa corrección (Burge 2003a, p. 282). Por este motivo, si un sujeto como Adam posee la habilidad de aplicar correctamente el término “sofá” entonces, de acuerdo con las pautas del externismo social, este sujeto poseerá el concepto correspondiente. En suma, de acuerdo con McLaughlin y Tye, el protagonista del ejemplo de Brown es un hablante que (i) posee un concepto de género no natural, (ii) es agnóstico sobre sus condiciones de aplicación y, además, (iii) no pertenece a ninguna comunidad lingüística. Pero esto muestra, justamente, que no podemos legitimar el establecimiento *a priori* de la premisa (RBr 2), porque la propia situación de Adam ilustra, precisamente, que la implicación afirmada por esta premisa no es verdadera.

Al margen del argumento de McLaughlin y Tye cabe realizar otra objeción a la presunción de Brown de que la premisa (RBr 2) puede ser establecida *a priori* a partir del externismo, la cual es, a su vez, una teoría conocida *a priori*. Tal y como se ha visto, las pautas del externismo social destacan los marcos de normatividad lingüística establecidos por otros hablantes como el elemento externo imprescindible en este sentido, puesto que son necesarios para la individuación de los conceptos tanto de género natural como de género no natural. Ahora bien, ya dijimos que resulta bastante dudoso que un partidario del externismo social (o natural) argumente esta tesis sobre

una base puramente apriorística, en el sentido en que la *reductio* de Brown requiere. Así, Burge en ningún momento muestra una actitud agnóstica sobre la existencia del mundo externo cuando formula sus tesis sobre la dependencia externista de nuestros conceptos, sino que *da por hecho* que existe el mundo externo (que incluye a los demás hablantes) y encuentra refuerzo para sus argumentos en las intuiciones de hablantes que también lo dan por hecho.²¹⁸ Por este motivo, no resulta en absoluto paradójico que un partidario del externismo social pueda establecer que no vive solo en el mundo a partir del hecho de que posee conceptos. La razón es que para proclamarse partidario de esta tesis no lo ha podido poner en duda. El externismo social, por tanto, cuenta con la existencia de otros hablantes distintos de nosotros mismos cuando apela a aquellas intuiciones que rigen nuestras atribuciones de pensamientos a terceros, por lo que resulta ciertamente dudoso que la premisa (RBr 2) pueda conservar un carácter puramente apriorístico como Brown sostiene.

Antes de dar por finalizada esta subsección es conveniente recapitular los aspectos más interesantes de las objeciones compatibilistas a las distintas versiones del argumento de la consecuencia. En primer lugar, hemos visto que una de las alegaciones más frecuentes consiste en que la noción de aprioricidad que se maneja en este argumento es sumamente restrictiva. De acuerdo con McKinsey, Boghossian y Brown, el conocimiento *a priori* es aquel conocimiento que puede lograrse sin necesidad de investigaciones y/o asunciones empíricas de ninguna clase. Pero, por una parte, es muy cuestionable pensar que las intuiciones que respaldan al externismo no cuentan con ninguna creencia empírica de fondo y, por otra parte, cabe rechazar que la accesibilidad fiable e inmediata que poseemos con respecto a nuestros propios pensamientos pueda concebirse en términos de aprioricidad. Además, aunque admitiéramos que las tesis externistas pueden establecerse *a priori* en algún sentido y aceptáramos describir el acceso privilegiado como acceso *a priori* a los propios pensamientos, aún cabría cuestionar, como indica Nuccetelli, que estemos usando la noción de aprioricidad en el mismo sentido en ambos casos.

²¹⁸ Algunas de las palabras más explícitas de Burge en este sentido son las siguientes: “El interés antiescéptico del anti-individualismo proviene, en parte, de la opinión de que esta tesis puede ser conocida mediante reflexión *a priori* [...] Pero el estatus preciso del anti-individualismo – o de varios aspectos de él – no resulta obvio. Parece comenzar a partir de la reflexión sobre la relación entre estados mentales y un entorno físico o social que, simplemente, se asume como existente. Además, ciertas características del anti-individualismo – tales como nuestras reflexiones sobre el papel de la evolución o las diferencias entre las distintas percepciones y los sistemas perceptivos de los individuos – parecen deberse a conocimientos empíricos de fondo” (Burge 2003, p. 339).

En segundo lugar, muchos autores compatibilistas coinciden en señalar que las distintas versiones del argumento de la consecuencia están basadas en una comprensión inadecuada de las tesis externistas. Así, la premisa (McK 2) del argumento de McKinsey describe el externismo como una tesis que establece una relación de dependencia lógica o conceptual (no metafísica) entre nuestros pensamientos y los factores externos que determinan sus contenidos. Ahora bien, de acuerdo con lo dicho en las páginas anteriores, creemos que el argumento que McKinsey formula para excluir la comprensión metafísica del externismo es una reducción al absurdo de una tesis *deliberadamente débil*. En cuanto al argumento de Boghossian, la mayoría de las réplicas compatibilistas se han dirigido hacia la imposibilidad de establecer *a priori* la verdad del antecedente de la segunda premisa, a saber, que seamos capaces de conocer cuáles de entre nuestros conceptos son conceptos atómicos que pretenden ser aplicados a géneros naturales. El principal problema residía, según comprobamos, en las tensiones internas a la distinción entre conceptos atómicos y compuestos propuesta por Boghossian. En cuanto a la *reductio* de Brown, su gran virtud frente a los planteamientos de McKinsey y Boghossian es la voluntad explícita de su autora de atender lo más fielmente posible a los textos de Burge a la hora de formular sus premisas. Pero, por una parte, el caso que Brown empleaba para demostrar que la inserción en una comunidad lingüística es una condición de posesión conceptual (de acuerdo con el externismo social) contravenía las indicaciones del propio Burge en este sentido. Por otra parte, hemos alegado que este autor tampoco estaría de acuerdo con otro presupuesto fundamental de la *reductio* de Brown, a saber, que la verdad de las tesis externistas, así como su instanciación en casos concretos, puede establecerse *a priori*. De hecho, la propia formulación del anti-individualismo de Burge requiere contar de antemano con la existencia de un entorno externo y de otros hablantes que habitan en él, por lo que resulta cuestionable que esta tesis quede establecida puramente *a priori*, sin contar con asunciones empíricas.

2.2 Cuestionamiento del carácter paradójico de la conclusión

Las réplicas compatibilistas al argumento de la consecuencia que hemos examinado hasta el momento presentan un denominador común, a saber, todas ellas pretenden mostrar que la conclusión de este argumento no puede obtenerse *a priori* a partir de la conjunción del externismo y el acceso privilegiado. En este sentido, lo que se alega es o

bien que las premisas del argumento de la consecuencia se apoyan sobre una concepción errónea del externismo y/o del acceso privilegiado, o bien que la verdad del externismo en general (y sus instanciaciones en casos particulares) no puede establecerse *a priori* en el sentido de *a priori* que asumen McKinsey, Boghossian y Brown.

Sin embargo, algunos partidarios del compatibilismo han optado por una estrategia distinta (y en gran medida opuesta) a la hora de encarar este desafío incompatibilista. A lo largo de esta subsección nos ocuparemos de un planteamiento alternativo, cuya intuición de partida es que la posibilidad de obtener *a priori* la conclusión del argumento de la consecuencia no tiene un carácter intrínsecamente paradójico. De acuerdo con esta nueva iniciativa, y al contrario de lo que piensan McKinsey, Boghossian y Brown, la capacidad de conocer *a priori* hechos empíricos, tales como la existencia de los determinantes externos de los contenidos de nuestros pensamientos, constituye un resultado no sólo admisible sino sumamente importante. El compatibilismo se contempla como una vía teórica que podría inspirar la construcción de argumentos que nos permitieran refutar determinadas hipótesis escépticas.

Para examinar esta estrategia, vamos a analizar con cierto detalle uno de estos escenarios escépticos, a saber, el escenario de los cerebros en cubetas presentado por Putnam en (1981). A continuación, primeramente tomaremos en consideración un argumento anti-escéptico, la prueba del mundo externo propuesta por Moore en (1939), cuya inviabilidad ha sido constatada de manera general. Nuestra pretensión es mostrar qué tiene de peculiar el argumento anti-escéptico derivado del compatibilismo frente al argumento de Moore y por qué los partidarios del primero consideran que nos permite esquivar las causas del fracaso del segundo. En segundo lugar, examinaremos la refutación que Putnam ofrece de su propia hipótesis escéptica de los cerebros en cubetas. En tercer y último lugar, analizaremos el empleo anti-escéptico del argumento de la consecuencia por parte de Sawyer y Warfield.

2.2.1 *La hipótesis escéptica de los cerebros en cubetas*

Atendiendo al orden de nuestros objetivos, comenzaremos explicando el diseño del escenario de los cerebros en cubetas, al cual nos referiremos de ahora en adelante como

“hipótesis o escenario BIV”, de acuerdo con las convenciones al uso.²¹⁹ El escenario que Putnam nos presenta es el siguiente. Supongamos que los cerebros de algunas personas han sido extraídos de sus cuerpos y colocados en una serie de cubetas. Estos cerebros así dispuestos están conectados mediante electrodos a una supercomputadora, cuya función es simular una corriente de experiencias para cada uno de estos cerebros. Mediante una serie de impulsos eléctricos se induce a estos cerebros la ilusión de estar viendo árboles, bebiendo agua, paseando por la calle, etc. El aspecto filosóficamente relevante de este escenario es exactamente el mismo que se presenta en la hipótesis del Genio Maligno de Descartes expuesta en la segunda sección del segundo capítulo de nuestro trabajo. En ambos casos se describe una situación en la que todas las experiencias de sus protagonistas son inducidas por un agente externo (el Genio o la supercomputadora), de manera que resultan subjetivamente indistinguibles de las experiencias que tendría una persona que estuviese, de hecho, viendo árboles, bebiendo agua o paseando por la calle. La pretensión de estos escenarios es aportar plausibilidad (al menos momentánea) a las inquietudes del escéptico sobre la existencia del mundo externo, ya que, *por lo que a cada uno de nosotros respecta*, es concebible que estemos ocupando la posición de los protagonistas de estos escenarios escépticos, esto es, que estemos viviendo en “mundos vacíos”.²²⁰ En suma, si fuéramos cerebros en cubetas, todo nos parecería exactamente igual a como nos parece actualmente porque, por hipótesis, tanto el mundo externo como la supercomputadora son capaces de generar un entramado de experiencias idéntico para nosotros.²²¹

²¹⁹ Del inglés, “Brains in a vat”, cerebros en una cubeta.

²²⁰ Emplearemos la denominación de “mundo vacío” para la clase de mundos externos postulados por las hipótesis escépticas. Aunque en el caso del escenario BIV el mundo externo contiene, al menos, los cerebros, las cubetas y la supercomputadora, la vacuidad de ese mundo refleja el hecho de que las experiencias del mundo que la supercomputadora induce a los cerebros no remiten a ese mundo externo, sino que son “vacías” en el sentido de que el cerebro no ve árboles reales, ni huele flores reales, etc.

²²¹ Otra manera de describir el reto filosófico que plantean los escenarios escépticos es apelar a la noción de “infradeterminación”. El origen de este concepto se sitúa en el seno de la filosofía de la ciencia. Se dice que dos teorías científicas están infradeterminadas por la evidencia si y sólo si, aunque estas teorías sean incompatibles entre sí, son compatibles con toda la evidencia disponible. En el caso de las hipótesis escépticas nos enfrentamos a una situación de infradeterminación generalizada y radical. De este modo, se ofrecen dos explicaciones alternativas y mutuamente incompatibles acerca de todas nuestras experiencias: o bien están causadas por elementos del mundo externo, o bien no hay un mundo externo tal y como lo conocemos y nuestra experiencia nos es inducida por medios artificiales. En cualquier caso, no podemos apelar a ningún tipo de evidencia empírica para optar por una alternativa u otra, puesto que, por hipótesis, tanto si experimentamos el mundo como si somos “engañados” por la supercomputadora, nuestras evidencias empíricas serán exactamente las mismas por lo que a nosotros respecta.

Una vez que hemos examinado la construcción de la hipótesis escéptica BIV y el problema filosófico al que da lugar, vamos a retomar el famoso argumento elaborado por G. E. Moore para refutar al escéptico e indicaremos las razones por las que suele considerarse que este intento es infructuoso.²²²

La estructuración canónica del argumento de Moore es la siguiente:²²³

PRUEBA DEL MUNDO EXTERNO DE MOORE

(Moo 1) Aquí hay dos manos.

(Moo 2) Si las manos existen, el mundo externo existe.

(Moo 3) El mundo externo existe.

Moore enuncia este argumento para extraer la conclusión de que no estamos viviendo en mundos vacíos como los que se describen en las hipótesis escépticas.²²⁴ Esto es, puesto que tengo dos manos, existe un mundo externo que contiene, al menos, mis dos manos. Como podemos formular argumentos semejantes con respecto a cualesquiera objetos de cuya existencia tengamos constancia (empírica), entonces, según Moore, el escéptico no puede sostener que vivimos radicalmente engañados en el sentido de que nuestra existencia se desarrolla, en realidad, en un mundo vacío.

Ahora bien, generalmente se admite que el argumento de Moore adolece de un defecto epistémico que socava su efectividad, ya que un verdadero escéptico no sólo pone en cuestión la verdad de la conclusión del argumento mooreano, sino también todas las justificaciones que podrían usarse para garantizar la primera premisa del

²²² Pérez Otero ha defendido que existen razones para considerar que argumentos como el de Moore no son completamente fallidos. A este respecto, el autor defiende que la elaboración de argumentos puede perseguir un fin “desenredador”, en el sentido de que podrían servir para hacer consciente al destinatario de que mantiene ciertas creencias y, además, para hacer explícitas las relaciones de consecuencia lógica que existen entre las mismas (2012, p. 61). Si aceptáramos que un argumento como el de Moore tiene éxito en su propósito desenredador, podríamos considerar que hay algo valioso en el mismo, a pesar de que existan destinatarios (típicamente, los escépticos) con respecto a los cuales el argumento carece de eficacia, entendida como fracaso a la hora de reforzar la creencia de éstos en la verdad de su conclusión. Por lo que a nosotros respecta, sólo tomaremos el argumento de Moore y el argumento de la consecuencia en tanto que formulados ante esta clase de destinatarios escépticos, es decir, para resolver la clase de dudas que plantean estos sujetos. Debido a esto, nuestras consideraciones serán independientes del diagnóstico de Pérez Otero en relación a otra clase de propósitos o destinatarios (un diagnóstico muy plausible, en nuestra opinión).

²²³ Cf. Davies (1998), (2000), (2003b); Wright (2002) y Pryor (2004). Esta estructuración recoge el argumento expuesto por Moore en el siguiente fragmento: “[P]uedo probar, por ejemplo, que existen dos manos humanas. ¿Cómo? Levantando mis dos manos y diciendo, mientras hago un cierto gesto con la mano derecha, ‘aquí hay una mano’, y añadiendo, mientras hago cierto gesto con la mano izquierda, ‘aquí hay otra’. Y si, por hacer esto, he probado *ipso facto* la existencia de las cosas externas, verán ustedes que ahora puedo hacer lo mismo de incontables maneras: no hay necesidad de multiplicar los ejemplos” (Moore 1939, pp. 145-146).

²²⁴ El enemigo filosófico de Moore en la formulación de su argumento es, en realidad, el idealismo más bien que el escepticismo. Sin embargo, en la medida en que tanto el idealista como el escéptico niegan que exista el mundo externo en el sentido en que el argumento de Moore pretende establecer, en lo que sigue consideraremos que dicho razonamiento tiene un propósito fundamentalmente anti-escéptico.

argumento, (Moo 1). Para entender esto, tengamos en cuenta que la función de los escenarios escépticos como BIV es mostrar que, de darse la situación que describen, quedaría en suspenso el valor justificativo de todas nuestras garantías empíricas. Por esta razón, si el escéptico cree, apoyándose en el escenario BIV, que la conclusión de Moore es falsa (no es verdad que el mundo externo existe), esta misma creencia le impediría conceder que las evidencias empíricas aducidas a favor de las premisas de su argumento (como percibir dos manos) constituyen un indicio de la verdad de estas premisas.²²⁵

2.2.2 *Los argumentos de Putnam contra la hipótesis escéptica de los cerebros en cubetas*

Tras haber examinado qué problemas presenta la prueba del mundo externo mooreana, vamos a atender a la propia posición de Putnam en (1981), quien tras formular la hipótesis BIV ofrece un argumento con el objetivo de demostrar que si fuéramos los protagonistas de dicha hipótesis no podríamos pensar o decir con verdad que lo somos y que esta conclusión podría emplearse con propósitos anti-escépticos.

Este autor sostiene que la lectura escéptica del escenario BIV atenta contra las *precondiciones necesarias del vínculo referencial*, ya se establezca éste entre nuestros términos lingüísticos y las entidades externas que designan, o bien entre nuestras imágenes mentales y los aspectos del mundo que representan (Putnam 1981, p. 29). En concreto, Putnam se enfrenta a las *teorías mágicas de la referencia* que conciben la relación referencial (lingüística o mental) como una propiedad intrínseca y necesaria de las representaciones. Estas teorías vendrían a sostener que las representaciones se refieren a lo representado por sí mismas e independientemente de que la representación

²²⁵ No obstante, hay que decir que Moore era consciente de este problema. En las páginas que siguen a la formulación de su argumento reconoce que es posible que muchos “filósofos” se sientan insatisfechos por la solución aportada (Moore 1939, pp. 149 y ss.). Las razones de la insatisfacción son dos. La primera, que no ha ofrecido ninguna prueba para demostrar la verdad de proposiciones empíricas como la que figura en la primera premisa de su argumento. La defensa de Moore es la siguiente: “¿cómo pruebo ahora ‘aquí hay una mano y aquí hay otra’? No creo que pueda hacerlo. Para ello, necesitaría probar, como Descartes señaló, que no estoy soñando ahora. Pero ¿cómo podría probar eso ahora? Sin duda, carezco de razones conclusivas para aseverar que ahora no estoy durmiendo” (Moore 1939, p. 149). El segundo motivo de insatisfacción es que, si no ha ofrecido ninguna prueba para demostrar la premisa (Moo 1), es imposible *conocer* la proposición que expresa dicha premisa. Ahora bien, Moore sostiene que la opinión de que o bien puedo ofrecer una prueba de la verdad de una proposición y entonces la conozco o bien sólo cabe creer en su verdad como una cuestión de fe es una disyuntiva propia de “filósofos”.

esté vinculada causalmente con lo representado.²²⁶ Por el contrario, Putnam defiende que la capacidad de remitir referencialmente a otros objetos es una propiedad extrínseca y contingente de las representaciones, una propiedad que depende del modo en que dichas representaciones son causadas. De este modo, si un hablante puede pensar en y referirse a los objetos externos se debe a que ha interactuado causalmente con esta clase de entidades.

A partir de estas afirmaciones de Putnam podemos extraer dos consecuencias. La primera de ellas es que una oración como, por ejemplo, “ahí hay un árbol”, proferida por un cerebro inserto en una cubeta, no expresa, según este autor, el mismo pensamiento que nosotros expresaríamos mediante una proferencia homófona. En efecto, dado que los árboles no han figurado, por hipótesis, en la historia causal del cerebro en la cubeta, no podemos decir que el cerebro esté pensando en árboles cuando emite esta proferencia, mientras que nosotros sí lo hacemos. La segunda consecuencia, de sumo interés para nosotros, es que, de acuerdo con Putnam, no podemos ser cerebros en una cubeta o, al menos, que si lo fuéramos, no podríamos decir o pensar con verdad que lo somos.²²⁷ Vamos a ofrecer la siguiente reconstrucción del argumento que Putnam ofrece para reforzar esta segunda consecuencia:²²⁸

²²⁶ Para Putnam si unas criaturas como nosotros nunca hubieran visto un árbol y se hiciera llegar hasta ellos una fotografía de un árbol, no habría motivo para decir que la imagen mental que se forjan dichas criaturas a la vista de la fotografía representa *per se* esa clase de objetos (aunque ellos no lo sepan). Según Putnam, describir así su imagen mental presupone nuestro punto de vista, a saber, el punto de vista de seres capaces de tener imágenes mentales que remiten a árboles reales, precisamente, por haber interactuado con esta clase de objetos. Cf. Putnam (1981), p. 17.

²²⁷ Preferimos expresar de esta manera las pretensiones del argumento de Putnam ya que su alcance es una cuestión controvertida y sus indicaciones al respecto no son muy claras. Por un lado, este autor sostiene que este argumento demuestra que si fuéramos cerebros en una cubeta, no podríamos decir o pensar que lo somos (Putnam 1981, p. 20). Sin embargo, en otras ocasiones Putnam otorga un mayor potencial a su argumento y llega a decir lo siguiente: “podemos ofrecer un argumento que demuestre que no somos cerebros en una cubeta” (Putnam 1981, p. 21).

²²⁸ Esta reconstrucción se corresponde con el siguiente fragmento de Putnam: “si su ‘mundo posible’ [el mundo posible de un individuo] es el mundo real, y somos realmente cerebros en una cubeta, entonces lo que queremos decir con ‘somos cerebros en una cubeta’ es que somos cerebros en una cubeta-aparente-en-la-imagen o algo de esta índole (si es que queremos decir algo). Pero parte de la hipótesis de que somos cerebros en una cubeta es que no somos cerebros en una cubeta-aparente-en-la-imagen (es decir, lo que estamos ‘alucinando’ no es que somos cerebros en una cubeta). Así pues, si somos cerebros en una cubeta, entonces la oración ‘somos cerebros en una cubeta’ afirma algo falso (si es que afirma algo). En resumidas cuentas, si somos cerebros en una cubeta, entonces ‘somos cerebros en una cubeta’ es falso. Por lo tanto es (necesariamente) falso” (Putnam 1981, pp. 27-28).

0. O bien soy un cerebro en una cubeta o bien no lo soy.

1. Supongamos que no soy un cerebro en una cubeta:

1.1 Si no soy un cerebro en una cubeta, la oración “soy un cerebro en una cubeta” es una oración del castellano.

1.2 Si la oración “soy un cerebro en una cubeta” es una oración del castellano, entonces es verdadera si y sólo si soy un cerebro en una cubeta.

1.3 Por hipótesis, no soy un cerebro en una cubeta.

1.4 Por tanto, la oración del castellano “no soy un cerebro en una cubeta” es falsa.

2. Supongamos que soy un cerebro en una cubeta:

2.1 Si soy un cerebro en una cubeta, la oración “soy un cerebro en una cubeta” es una oración del lenguaje BIV.

2.2 Si la oración “soy un cerebro en una cubeta” es una oración del lenguaje BIV, entonces es verdadera si y sólo si soy un cerebro-en-la-imagen en una cubeta-en-la-imagen.²²⁹

2.3 Por hipótesis, no soy un cerebro-en-la-imagen en una cubeta-en-la-imagen.

2.4 Por tanto, la oración del lenguaje BIV “no soy un cerebro en una cubeta” es falsa.

Por (1.1)-(1.4), si no soy un cerebro en una cubeta y profiero la oración castellana “soy un cerebro en una cubeta” digo algo falso.

Por (2.1)-(2.4), si soy un cerebro en una cubeta y profiero la oración del lenguaje BIV “soy un cerebro en una cubeta” digo algo falso.

La conclusión que Putnam pretende extraer de su argumento es que tanto si la oración “somos cerebros en una cubeta” es proferida por un hablante del mundo actual como si es proferida por un cerebro en una cubeta, se trata de una oración falsa. Desde luego, en el primer caso sería una oración indudablemente falsa, porque en este mundo dicho hablante no es un cerebro en una cubeta, sino una persona que interacciona con elementos del mundo externo. Por otro lado, Putnam nos dice que si la oración “somos

²²⁹ Putnam sostiene que no es seguro que las preferencias del lenguaje BIV tengan siquiera condiciones de verdad, pero afirma que, en caso de tenerlas, éstas serían idénticas a las imágenes generadas por la supercomputadora o a los impulsos eléctricos que ésta efectúa. En nuestra reconstrucción optaremos por otorgar condiciones de verdad de la primera clase a las oraciones del lenguaje BIV. De esta manera una oración del lenguaje BIV como “el cielo es azul” sería verdadera si y sólo si el cielo-en-la-imagen es azul-en-la-imagen.

cerebros en una cubeta” fuese proferida por un cerebro en una cubeta, *también sería una oración falsa*. La razón es que las únicas condiciones de verdad que cabe otorgar a las oraciones proferidas por los cerebros en las cubetas son las imágenes fabricadas por la supercomputadora. Por ejemplo, si el cerebro profiriera la oración “hoy brilla el sol” y formara parte de la alucinación inducida que el sol-en-la-imagen brilla-en-la-imagen, dicha oración sería verdadera y, si no se dieran tales imágenes, sería falsa. Con respecto a la oración “somos cerebros en una cubeta” ocurre lo siguiente. La situación objetiva de ser un cerebro en una cubeta es algo que, por principio, no puede ser experimentado o representado mediante imágenes inducidas por una hipotética supercomputadora. La razón es que dicha situación rebasa las condiciones de la representación para un cerebro en una cubeta establecidas en el externismo de Putnam, unas condiciones que establecen que un sujeto sólo puede representar aquello que pertenece al ámbito de sus interacciones causales. Si yo fuera un cerebro en una cubeta no podría entablar contactos causales con la cubeta que me contiene, sino sólo con cubetas-en-la-imagen y, por ello, no sería capaz de pensar en o referirme a dicha cubeta, sino sólo a esas cubetas-en-la-imagen. A partir de esto, se entiende que la oración “soy un cerebro en una cubeta” proferida por un cerebro en una cubeta no puede ser verdadera, puesto que los términos involucrados en ella no tienen como referente ni el cerebro que él mismo es ni la cubeta que le contiene. Si cabe otorgar condiciones de verdad a las oraciones de este sujeto sólo podrán serlo las imágenes que la supercomputadora le induce y, puesto que entre estas imágenes no se encuentra la de ser un cerebro en una cubeta, la oración “soy un cerebro en una cubeta” dicha por un cerebro en una cubeta es una oración falsa. Por ello, concluye Putnam, la oración “soy un cerebro en una cubeta” será falsa *tanto* en el mundo actual, donde no soy un cerebro en una cubeta, *como* en el mundo posible donde sí lo soy.²³⁰

²³⁰ Sin embargo, tal y como señala Brueckner en (1986, p. 152) la conclusión que Putnam extrae de este argumento no puede ser la necesaria falsedad de la oración “soy un cerebro en una cubeta”. La razón es que no estamos sometiendo a evaluación contrafáctica la *misma* oración “soy un cerebro en una cubeta” que profiero en el mundo actual, ya que uno de los requisitos de esta clase de evaluaciones es que la semántica de la oración en cuestión permanezca inalterada, pero esto es algo que, por hipótesis, no ocurre. Desde el momento en que se toma en consideración un entorno contrafáctico en el que soy un cerebro en una cubeta, el lenguaje en el que se enmarca mi preferencia ya no es el castellano, dado que han variado radicalmente las propiedades semánticas (significado, referencia y condiciones de verdad) de este lenguaje que, por otra parte, es fonética y sintácticamente idéntico al castellano. Podemos explicarlo de otra manera. Decir que una oración es necesariamente falsa supone mantener que *lo dicho* por esa oración no es el caso en ningún mundo posible. Pero, como vemos, lo dicho por esa oración cuando es proferida en castellano por el hablante actual y lo dicho por una oración homófona que pertenece al lenguaje BIV no es lo mismo, aun cuando cada una de estas oraciones sea falsa en los dos lenguajes respectivos.

Dicho esto, vamos a realizar algunas observaciones sobre el argumento de Putnam, partiendo de la reconstrucción del mismo que hemos ofrecido. En primer lugar, podríamos preguntarnos si nos autoriza a concluir puramente *a priori* que no somos cerebros en cubetas. Ya hemos visto el interés que tiene esta cuestión, puesto que si alguna de las premisas del razonamiento de Putnam ha sido establecida a partir de la experiencia, entonces se estaría empleando un apoyo empírico cuya fiabilidad, por principio, está puesta en cuestión por el escéptico. A este respecto, el propio Putnam afirma que su razonamiento ha sido establecido *a priori* aunque no en un sentido absoluto (Putnam 1981, p. 29), lo cual significa, atendiendo a sus propias pautas, que su metodología toma en consideración supuestos falibles y empíricos, tales como la tesis de que la mente no tiene otro acceso que el acceso causal a las cosas externas. De ser esto cierto, la eficacia anti-escéptica del argumento se vería seriamente afectada.

En segundo lugar, en (2003b) Grimaltós ha señalado que aun cuando el razonamiento de Putnam permitiera concluir que no podemos decir o pensar con verdad la oración “soy un cerebro en una cubeta”, dicho razonamiento no ofrece ninguna conclusión directamente anti-escéptica que nos sirva para demostrar que, en realidad, no somos cerebros en cubetas. Esto es, la reconstrucción del razonamiento que Putnam lleva a cabo en (1981, pp. 27-28) concluye que la oración del castellano “soy un cerebro en una cubeta” es falsa y que una oración homófona perteneciente al lenguaje BIV también lo sería, pero no que yo no sea un cerebro en una cubeta. Desde luego, si Putnam pretende haber ofrecido un argumento que permita extraer de manera directa esta última conclusión, entonces dicho argumento no sería idéntico al que hemos formulado en las páginas anteriores. Por este motivo, ofreceremos el esquema de un hipotético argumento anti-escéptico, susceptible de ser extraído a partir de las tesis de Putnam, que no coincide con el expuesto por este autor en el fragmento citado y que, además, guarda cierto parecido con el argumento de la consecuencia:

(Put 1) El contenido de mi pensamiento es que no soy un cerebro en una cubeta.

(Put 2) Las condiciones para establecer un vínculo representacional son de carácter externista y causal.

(Put 3) Por (Put 2), si yo fuera un cerebro en una cubeta entonces no podría pensar que soy un cerebro en una cubeta.

(Put 4) Por (Put 1) y (Put 3), no soy un cerebro en una cubeta.

En nuestra opinión Grimaltós está en lo correcto al observar que resulta muy aventurado extraer (Put 4), la conclusión anti-escéptica, de este razonamiento. Para empezar, cabe alegar que el propósito del escéptico al enunciar el escenario BIV es mostrar que, por todo lo que sabe el sujeto que enuncia el argumento, él no tiene garantías para establecer la verdad de la premisa (Put 1) más bien que su negación. Así, aunque aceptáramos la verdad del externismo recogida en la premisa (Put 2) y la implicación establecida en (Put 3), el sujeto tendría que conocer de antemano la verdad de (Put 4) para poder aceptar la premisa (Put 1). O, dicho de otro modo, uno no puede saber, si se toma en serio la hipótesis escéptica, si su lenguaje es el castellano o si es el lenguaje BIV. Es decir, uno no puede conocer el supuesto del que depende la verdad de la premisa (Put 1), ya que si no se conoce qué lenguaje está hablando, tampoco puede saber si el pensamiento que está expresando es el pensamiento de que es un cerebro en una cubeta o si es el pensamiento de que es un cerebro-en-la-imagen en una-cubeta-en-la-imagen.²³¹

Las consideraciones precedentes sobre la hipótesis BIV y sobre los recursos de Putnam para refutar al escéptico nos permiten apreciar que la posición de este autor no es todo lo eficaz que sería deseable, ya que él mismo admite que sus argumentos no tienen un carácter puramente *a priori*, como se requiere de las refutaciones anti-escépticas, y, además, tampoco sirven para demostrar que no somos cerebros en cubetas, sino, a lo sumo, que tanto si soy un cerebro en una cubeta como si no lo soy mis preferencias de la oración “soy un cerebro en una cubeta” serán preferencias de una oración falsa.

2.2.3 *La posibilidad de un uso anti-escéptico del argumento de la consecuencia: Sawyer y Warfield*

Una vez examinados los argumentos de Putnam contra la hipótesis BIV, nos ocuparemos en esta subsección de las iniciativas de Sawyer en (1998) y Warfield en (1998), quienes continúan la línea ensayada por Putnam en (1981) en cuanto que pretenden dar un uso anti-escéptico a las tesis externistas. Los trabajos de estos autores

²³¹ El argumento escéptico contra el establecimiento de (Put 1) tendría la forma siguiente:

(Esc 1) No sé si soy un cerebro en una cubeta.

(Esc 2) Si no sé si soy un cerebro en una cubeta no sé si estoy pensando que no soy un cerebro en una cubeta (más bien que no soy un cerebro-en-la-imagen en una cubeta-en-la imagen).

(Esc 3) No sé si estoy pensando que no soy un cerebro en una cubeta (más bien que no soy un cerebro-en-la-imagen en una cubeta-en-la imagen).

presentan una alegación común, a saber, que la conclusión obtenida *a priori* en cualquiera de las versiones del argumento de la consecuencia no es de ninguna manera una conclusión inadmisibles. De acuerdo con ambos autores, la capacidad de lograr conocimiento *a priori* de hechos empíricos sustantivos resulta ser, en efecto, una consecuencia natural y directa del compatibilismo, pero esto no menoscaba en absoluto la plausibilidad de esta última tesis. Sin embargo, es preciso decir que aunque esta conclusión básica se encuentra presente en los trabajos de ambos autores, las pretensiones y supuestos de uno y otro son un tanto distintos.

En lo que respecta a Sawyer, esta autora realiza un énfasis mucho menor que Warfield en el presunto potencial anti-escéptico del argumento de la consecuencia.²³² Por lo general, se limita a argumentar que, aunque a primera vista la posibilidad de obtener *a priori* la conclusión de la *reductio* puede resultar paradójica, ello no tiene un carácter intrínsecamente inaceptable. El núcleo de la argumentación de Sawyer es la intuición de que la explicación externista de los mecanismos de adquisición conceptual es mucho más plausible que la visión internista al respecto. En este sentido, de aceptar el externismo, tendremos que aceptar que la adquisición de conceptos requiere la interacción causal con un entorno físico o con un entorno lingüístico determinado. Si renunciamos a esta tesis, tendremos que aceptar que un sujeto puede poseer conceptos sin necesidad de establecer contacto causal con un entorno extra-mental (Sawyer 1998, p. 531). A partir de esto, indica nuestra autora, cualquiera que acepte la verdad de las tesis externistas estaría legitimado para plantear el siguiente argumento, similar al argumento de la consecuencia:

(Saw 1) Tengo pensamientos sobre el agua.

(Saw 2) Si tengo pensamientos sobre el agua entonces E.

(Saw 3) Por tanto, E.

En este caso “E” abrevia la siguiente disyunción: o bien vivo en un entorno que contiene agua, o bien estoy inserto en una comunidad de hablantes de quienes he aprendido el uso del término “agua”. Según nos dice Sawyer, no existe nada paradójico en el hecho de que un agente pueda derivar *a priori* conclusiones como E, siempre y cuando E recoja hechos empíricos que el agente debería poder conocer, además, por medios empíricos (Sawyer 1998, p. 532). De este modo, lo que nos proporciona el

²³² Así, Warfield comienza su artículo con una declaración explícita acerca del potencial anti-escéptico del compatibilismo y a lo largo de su trabajo vuelve a menudo sobre este asunto. Sin embargo, Sawyer se refiere a ello en una sola ocasión y casi de pasada en (1998), p. 531.

compatibilismo es *otra ruta epistémica* (además de la empírica) para alcanzar *a priori* el conocimiento de hechos contingentes como los mencionados en la conclusión de este argumento.

Ahora bien, en nuestra opinión la posición de Sawyer ha de atender a dos dificultades. La primera de ellas es que su comprensión del externismo, en tanto que una tesis sobre las condiciones de *adquisición* de nuestros conceptos y contenidos, no es la más usual. Según indicamos en la introducción de este trabajo, el internismo y el externismo psicológicos son concepciones enfrentadas en lo que respecta a las condiciones de individuación (y no de adquisición) de nuestros contenidos mentales.²³³ Esto es, un internista no tendría por qué rechazar que la adquisición de conceptos procede por vías causales, ya que lo único que tiene que defender en tanto que internista es que, independientemente de cómo se hayan adquirido sus conceptos, si dos hablantes son idénticos en todos sus aspectos intrínsecos (fenomenológicos, funcionales, etc.), entonces estos hablantes poseerán idénticos conceptos. O, lo que es igual, si se produjera una variación en los elementos externos a un hablante que dejara inalteradas sus propiedades internas, dicha variación no afectaría a la identidad de los conceptos incluidos en el repertorio del hablante.²³⁴ Si esto es así, Sawyer está en un error cuando afirma lo siguiente: “[e]n la imagen que nos presenta el individualismo, cabe imaginar a un sujeto equipado con ciertos conceptos antes de su exposición al mundo” (Sawyer 1998, p. 531). Según dejan entrever estas afirmaciones, nuestra autora piensa que el internista está comprometido con una tesis innatista sobre la adquisición conceptual. Pero ya hemos explicado que estas opciones teóricas son independientes.

En segundo lugar, estimamos que el intento de utilizar el argumento anterior como un recurso anti-escéptico está abocado al fracaso. Las consideraciones realizadas al hilo del trabajo de Putnam dejan ver que resulta muy cuestionable que un escéptico acepte las justificaciones de Sawyer para establecer la premisa (Saw 2). De acuerdo con esta autora, el núcleo de tales justificaciones viene dado por la intuición externista de que sólo puedo tener el concepto de agua si he interactuado causalmente con esta sustancia o con una comunidad lingüística donde se me ha enseñado el uso del término que expresa dicho concepto. Es cierto que esta intuición goza de gran verosimilitud, pero el problema es que si nos situamos en el espacio dialéctico de la argumentación anti-escéptica, Sawyer no puede ofrecer ninguna garantía (sin cometer petición de

²³³ Cf. Wilson (2003), pp. 256-257; Davies (2000), p. 322 y Moya (2004), pp. 153-154.

²³⁴ Nos detendremos más en este punto en la primera sección del quinto capítulo de este trabajo.

principio) para asegurar que poseo el concepto de agua y no sólo que me parece que lo poseo. Esto es, lo que el escéptico cuestiona es que uno pueda asegurar que ha interactuado causalmente con un entorno físico o lingüístico dado que, por todo lo que le parece, podría estar sólo “alucinando” dicha interacción. De este modo, si yo no sé si estoy alucinando, entonces tampoco sé si mis presuntas interacciones causales con el mundo son reales o alucinadas. Además, el hecho de que Sawyer reconozca que sólo aquellos hablantes que puedan saber por medios empíricos que existen los condicionantes externos de sus conceptos son, por ello, capaces de extraer *a priori* la conclusión (Saw 3), pone las cosas fáciles al escéptico. En la medida en que éste formula hipótesis que suspenden la fiabilidad de nuestras vías sensoriales de acceso al mundo, nadie que se tome en serio esta clase de hipótesis podría saber por esas vías que existen tales condicionantes contingentes y, por tanto, tampoco podría saberlo por medios apriorísticos.

En lo que respecta a Warfield, su posición es similar a la de Sawyer aunque, como hemos indicado, las pretensiones anti-escépticas de su trabajo son mucho más explícitas. En concreto, este autor pretende combatir el establecimiento de la primera premisa de un hipotético argumento escéptico mediante la capacidad de conocer hechos empíricos *a priori* que, según se alega, nos proporcionaría el compatibilismo. El argumento escéptico que Warfield toma en consideración tiene la siguiente forma:

(Esc-W 1) No sé si soy un cerebro en una cubeta en un mundo vacío.

(Esc-W 2) Si no sé si soy un cerebro en una cubeta en un mundo vacío no sé si actualmente bebo agua.

(Esc-W 3) Por tanto, no sé si actualmente bebo agua.

De acuerdo con Warfield, el compatibilismo permitiría demostrar la falsedad de la primera premisa (Esc-W 1) y concluir que no somos cerebros en cubetas. Su argumento es el siguiente:

(War 1) Pienso que el agua es húmeda.

(War 2) Ningún cerebro en una cubeta en un mundo vacío piensa que el agua es húmeda.

(War 3) No soy un cerebro en una cubeta en un mundo vacío (o al menos no siempre lo he sido).²³⁵

²³⁵ La salvedad incluida entre paréntesis indica que el alcance anti-escéptico del argumento de Warfield es más reducido de lo que inicialmente podría pensarse. Dicha salvedad responde al hecho de que, según Warfield, la posesión de pensamientos acerca del agua requiere que *en algún momento* un hablante haya

Según Warfield, la mayor fortaleza de este argumento anti-escéptico reside en el hecho de que, al contrario que la prueba del mundo externo mooreana, no comete petición de principio (Warfield 1998, p. 140). Esto es, mientras que Moore sólo puede establecer la primera premisa de su prueba a partir de justificaciones empíricas que, por definición, están desactivadas en las hipótesis escépticas, las premisas del argumento de Warfield pueden establecerse sin contar con justificaciones de esta naturaleza.

No obstante, el argumento de Warfield merece un diagnóstico muy similar al que hemos ofrecido con respecto al de Sawyer. En este sentido, si aceptamos la individuación externista de nuestros conceptos, sólo podemos establecer la premisa (War 1) al precio de cometer petición de principio, al igual que ocurría con la prueba del mundo externo mooreana. Esto es, si el hecho de que yo piense que el agua es húmeda (y no que sólo me lo parezca) depende de que yo sea un ser humano (y no un cerebro en una cubeta) y lo que alega el escéptico es que la situación en la que yo soy un ser humano me resulta completamente indistinguible de aquella en la que soy un cerebro en una cubeta, entonces sólo puedo establecer que pienso que el agua es húmeda si asumo de antemano la falsedad de la tesis escéptica. Dicho de otro modo, a partir de la aceptación del externismo, un escéptico podría formular el siguiente argumento con el objetivo de impedir el establecimiento de la premisa (War 1) del razonamiento anti-escéptico de Warfield:

(Esc 1) No sé si soy un cerebro en una cubeta.

(Esc 2) Si no sé si soy un cerebro en una cubeta, entonces no sé si estoy pensando que no soy un cerebro en una cubeta (más bien que no soy un cerebro-en-la-imagen en una cubeta-en-la imagen).

(Esc 3) No sé si estoy pensando que no soy un cerebro en una cubeta (más bien que no soy un cerebro-en-la-imagen en una cubeta-en-la imagen).

La justificación que ofrece el escéptico para aseverar la premisa (Esc 2) consiste en que una vez aceptadas las tesis externistas, existe una dependencia entre mi situación externa y la identidad de mis pensamientos. De este modo, si soy un ser humano inserto en un entorno no vacío tendré el pensamiento de que no soy un cerebro en una cubeta, pero si lo fuese, mi pensamiento sería que no soy un cerebro-en-la-imagen en una

estado inserto en un entorno que contiene agua. Sin embargo, podría ocurrir que, inicialmente, el individuo que formula el argumento haya vivido de ese modo, pero que, más tarde, su cerebro haya sido extirpado de su cuerpo e insertado en una cubeta sin que el individuo lo advierta. De este modo, si un hablante puede saber que tiene pensamientos acerca del agua, ello sólo demostraría que no siempre ha sido un cerebro en una cubeta, sin perjuicio de que pueda serlo en el preciso momento de formular el argumento.

cubeta-en-la-imagen. El problema es que la hipótesis escéptica cobra fuerza del hecho de que estas dos situaciones alternativas son, *por lo que a mí respecta*, completamente indistinguibles, lo cual significa que no puedo escoger justificadamente cuál de ellas es la verdadera. Pero si la capacidad de conocer qué es lo que pienso depende de saber si mi situación es la de un ser humano o la de un cerebro en un cubeta, que es lo que, justamente, no puede saberse de acuerdo con el escéptico, entonces éste puede alegar que en realidad no sé qué es lo que pienso y, por tanto, que la premisa (War 1) carece de justificación.

Llegados a este punto, conviene recapitular los aspectos más relevantes de nuestras consideraciones precedentes. Según hemos podido apreciar, no han faltado autores que han tomado el argumento de la consecuencia como un modelo de razonamiento anti-escéptico que, a diferencia del mooreano, escape a la acusación de petición de principio. El primero de estos intentos, como vimos, proviene de Putnam. Ahora bien, hemos alegado que su argumento sólo puede probar, como mucho, que la oración “soy un cerebro en una cubeta” proferida por un castellano hablante es falsa y que una preferencia homófona emitida por un cerebro en una cubeta también lo será. Sin embargo, como vimos, esto no permite extraer ninguna conclusión genuinamente anti-escéptica.

Los intentos de Sawyer y Warfield, por su parte, presentan dificultades muy similares. La afirmación de Sawyer de que sólo un agente que puede conocer por vías empíricas la existencia de determinantes conceptuales de carácter externo será capaz, a su vez, de lograr dicho conocimiento puramente *a priori*, pone en evidencia que si un sujeto no es capaz de saber que su conocimiento discurre por la primera ruta, tampoco podrá adquirir conocimiento *a priori*. Pero, por definición, un escéptico cuestiona lo primero y si no somos capaces de demostrar al escéptico que podemos garantizar la fiabilidad de la vía empírica, tampoco seremos capaces de lograr conocimiento *a priori* de hechos externos. Warfield, por su lado, confía en que la capacidad de conocer nuestros pensamientos conjugada con la dependencia que existe entre éstos y los condicionantes externos de sus contenidos pueda servirnos para deducir *a priori* la existencia de dichos condicionantes. El problema, como vimos, es que, dado un marco externista de individuación de contenidos, la identidad de nuestros pensamientos será distinta, en función de si es el mundo o la supercomputadora su determinante externo; pero, por lo que a nosotros respecta, todo nos parecería exactamente igual en cualquiera de los dos casos, lo que significa, en último término, que no seríamos capaces de

establecer la identidad de nuestros pensamientos sin apelar a factores empíricos cuya fiabilidad, por hipótesis, está puesta en duda.

2.3 *Petición de principio y transmisión de la garantía*

En esta última subsección vamos a examinar los trabajos de Wright y Davies sobre el problema de la transmisión de la garantía en argumentos lógicamente válidos. Estos autores han elegido un modo de enfrentarse al desafío planteado por el argumento de la consecuencia que, en parte, coincide con las propuestas de Warfield y Sawyer, pero que, en último término, extrae un diagnóstico opuesto al de estos autores. Así, al igual que Warfield y Sawyer, Wright y Davies son partidarios del externismo compatibilista y admiten que la verdad de las premisas del argumento de la consecuencia puede establecerse *a priori* (o al menos de manera no empírica). Pero, por otro lado, tanto uno como otro señalan que aquellos autores que formulan este argumento porque dudan de la verdad de su conclusión (porque son escépticos) no podrán convencerse de la verdad de ésta *como* una conclusión deducible de sus premisas, esto es, no podrán *obtener o ganar* conocimiento de su conclusión. Lo que ambos autores sostienen es, esencialmente, que el argumento de la consecuencia constituye una instancia de cierto tipo de razonamientos que presentan un defecto epistémico. En concreto, se trata de argumentos formalmente válidos, cuyas premisas, incluso, pueden ser verdaderas, pero que *no poseen un carácter convincente porque la verdad de su conclusión no queda garantizada por la verdad de sus premisas*. En suma, se trata de argumentos en los que se produce un fallo en la transmisión de la garantía de las premisas hacia la conclusión.

A fin de explicar las propuestas de estos autores con la mayor claridad posible es preceptivo realizar una serie de distinciones. De este modo, hay que diferenciar la cuestión de la *clausura sobre la implicación conocida* (o clausura epistémica), por un lado, y la cuestión de la *transmisión de la garantía*, por otro.²³⁶ El principio de clausura epistémica afirma que si un sujeto sabe que P y, además, sabe que P implica Q, entonces este sujeto sabe que Q. Esto significa que cuestionar la aplicación universal de este principio supone cuestionar que un sujeto sea capaz de conocer aquellas proposiciones (o algunas de ellas) implicadas por otra(s) proposición(es) que ya conoce. Desde luego, la cuestión de si este principio tiene una validez universal es algo que ha generado

²³⁶ Cf. Davies (1998), p. 325 y (2000), p. 393 y Wright (2000b), p. 16 y (2003), p. 57.

interesantes discusiones en epistemología. No obstante, no cabe aquí tratar en profundidad este asunto.²³⁷ Tan sólo recogeremos la problemática de la clausura epistémica para distinguirla de la cuestión de la transmisión de la garantía. En este último caso, si un argumento presenta un fallo en la transmisión de la garantía entonces el poder justificador de las garantías aducidas a favor de sus premisas no se transmite a la conclusión. En consecuencia, un hablante que formule un argumento de esta clase no ganará conocimiento de la conclusión y, en última instancia, no podrá quedar convencido de la verdad de ésta.

Las relaciones lógicas que existen entre la clausura epistémica y la transmisión de la garantía son las siguientes. En aquellos argumentos donde falla el principio de clausura, tampoco existe transmisión de la garantía de las premisas a la conclusión. De este modo, si sé que P y sé que P implica Q pero, a pesar de ello, no sé que Q, parece claro que cualesquiera que sean las garantías que tengo para saber que P y para saber que P implica Q no serán suficientes para saber que Q; el poder justificador de esas garantías no se transmitirá de las premisas a la conclusión, dado que no puedo conocer

²³⁷ Hay que decir, sin embargo, que si un autor consiguiera demostrar que este principio no se cumple en el argumento de la consecuencia, con ello abriría una posible vía para refutar al incompatibilista. Este es el caso de Dretske, uno de los autores que ha examinado con mayor profundidad el principio de clausura, quien defiende en (2003) que un individuo puede saber qué piensa (cómo se representa el mundo), puede considerar, además, que el externismo es una tesis correcta sobre la individuación conceptual y que, sin embargo, esto no le permitirá saber *que piensa* (sobre objetos de un mundo externo), aun cuando esta consecuencia esté implicada por la conjunción de las dos primeras tesis. De acuerdo con Dretske (2005, pp. 20 y ss.), existen “implicaciones pesadas”, esto es, proposiciones que sabemos que están implicadas por otras proposiciones que ya conocemos pero cuya verdad, sin embargo, no es cognoscible a partir de nuestro conocimiento de la verdad de estas últimas. Los ejemplos de implicaciones pesadas que Dretske nos proporciona son tres, a saber, que el mundo externo existe, que los otros seres humanos no son zombis carentes de mente y que el pasado es real. Es fácil ver en estas tres implicaciones las tesis que ponen en cuestión las tres variedades de escepticismo más desafiantes, es decir, el escepticismo acerca del mundo externo, acerca de las otras mentes y acerca de la existencia del pasado. Para Dretske el principio de clausura sólo tiene aplicación cuando la negación de la proposición implicada es una *alternativa relevante* a la proposición en cuestión. De esta manera, podemos pensar en una situación en la que trato de averiguar si una camiseta es una prenda de una determinada marca. Supongamos que si sé que esa prenda tiene un etiquetado determinado entonces sé que se trata de una prenda de esa marca. En este caso, saber si la camiseta posee el etiquetado en cuestión me permitiría excluir la alternativa relevante a la proposición de que la prenda es de la primera marca, a saber, que sea de otra distinta. Ahora bien, el problema que presentan las tres tesis escépticas es que no constituyen alternativas relevantes con respecto a nuestro conocimiento ordinario. Por tanto, Dretske desautoriza el uso que hace el escéptico del principio de clausura cuando sostiene que como no sé si soy un cerebro en una cubeta y como sé que tener manos implica que no soy un cerebro en una cubeta, entonces no sé si tengo manos. Dretske impide la maniobra escéptica al desautorizar la apelación al principio de clausura. Pero, además, este autor también cuestiona las refutaciones anti-escépticas “demasiado fáciles”, como aquella que se instancia en el argumento de la consecuencia. En este sentido, afirma Dretske, el modo (percepción o introspección) en que podemos conocer aquellas proposiciones (sé que tengo manos o sé que pienso que el agua es húmeda) que implicarían proposiciones pesadas no nos sirve para conocer la verdad de estas últimas. De acuerdo con este autor, “[l]a percepción, la manera más importante que tenemos de decir qué ocurre en el mundo externo, no nos dice que hay un mundo externo, y la introspección, nuestra manera de decir con autoridad qué ocurre en nuestras mentes, no nos dice que tengamos mentes” (Dretske 2003, p. 140).

esa conclusión en tanto que implicada por las premisas. Ahora bien, tanto Wright como Davies señalan que cabe pensar en argumentos donde la clausura se mantenga pero en los que no se produce una transmisión efectiva de la garantía desde las premisas a la conclusión.²³⁸ Esto sucederá si un hablante toma en consideración un argumento en el que acepta la verdad de las premisas, así como la relación de implicación que existe entre éstas y la conclusión y, sin embargo, las justificaciones que se aportan para garantizar las premisas son tales que no añaden garantía alguna a la conclusión, por lo que su poder justificador no se transmite en la ruta inferencial.

Además, tanto Wright como Davies admiten que existe un estrecho parentesco entre el fenómeno de petición de principio, por un lado, y el fallo en la transmisión de la garantía, por otro.²³⁹ Ahora bien, las consideraciones de ambos autores al respecto son un tanto distintas. De acuerdo con Wright, los argumentos que cometen petición de principio constituyen un subconjunto de los argumentos en los que no se transmite la garantía. Para Davies, por el contrario, el fenómeno de la petición de principio es un fenómeno dialéctico que tiene lugar sólo cuando un hablante formula un argumento (de ciertas características) para convencer *a un oyente* de la verdad de la conclusión. Sin embargo, el fallo en la transmisión de la garantía se produce cuando un hablante formula *para sí* un argumento (de ciertas características).²⁴⁰

Hechas estas distinciones de carácter introductorio, estructuraremos esta subsección en dos partes diferenciadas dedicadas a examinar y discutir las propuestas de cada uno de estos dos autores.

2.3.1 *Consideraciones de Wright sobre el fallo en la transmisión de la garantía*

En esta subsección vamos a tomar en consideración con cierto detenimiento las propuestas de Wright con el fin de examinar cuál es su tratamiento de los fenómenos de petición de principio y de fallo en la transmisión de la garantía. Lo primero que debemos señalar es que, de acuerdo con este autor, existen tres clases de fallo en la transmisión de la garantía. El primero de ellos, que denominaremos (FGT 1), se produce

²³⁸ Wright (2000b), p. 141 y Davies (1998), p. 326 y (2000), p. 393.

²³⁹ No han faltado autores que han señalado que ambos fenómenos son, esencialmente, uno y el mismo. De hecho, Pryor en (2001) sostiene que el nombre de “fallo de transmisión de la garantía” no es más que una nueva denominación que Wright y Davies proponen para el viejo problema de la petición de principio.

²⁴⁰ En concreto, lo que Davies afirma es que “el fenómeno del fallo en la transmisión de la garantía es el análogo, dentro del pensamiento de un único sujeto, del fenómeno dialéctico de cometer petición de principio” (Davies 2003a, p. 123). También cf. Davies (2000), p. 400.

en aquellos argumentos que cometen petición de principio en un sentido trivial, en cuanto que su conclusión figura explícitamente entre sus premisas.²⁴¹ El segundo tipo de petición de principio, o (FTG 2), tiene lugar en aquellos argumentos en los cuales la conclusión forma parte del cuerpo de informaciones que otorgan fuerza justificadora a las premisas. Dicho de otra manera, en los argumentos (FTG 2) ha de contarse de antemano con la verdad de la conclusión para que las evidencias aducidas a favor de las premisas tengan verdadero poder justificador. El tercer tipo de fallo en la transmisión de la garantía o (FTG 3), que Wright es reacio a calificar de petición de principio, es el que se presenta propiamente en el argumento de la consecuencia. En lo que sigue, explicitaremos los rasgos de los fallos de tipo (FTG 2) y (FTG 3) y veremos de qué modo este autor emplea sus conclusiones para desactivar el argumento incompatible.

Conviene comenzar examinando las características de los primeros argumentos que, de acuerdo con Wright, presentan un fallo en la transmisión de la garantía de carácter no trivial, es decir, un fallo de tipo (FTG 2). Para ello es recomendable atender a tres de los ejemplos que el autor propone a este respecto.²⁴²

ARGUMENTO GOL

(Gol 1) Se ha marcado un gol.

(Gol 2) Si se ha marcado un gol, se está jugando un partido.

(Gol 3) Por tanto, se está jugando un partido.

ARGUMENTO VOTO

(Voto 1) Juan ha marcado una X en una papeleta electoral.

(Voto 2) Juan acaba de votar.

(Voto 3) Se están celebrando elecciones.

ARGUMENTO MOORE

(Moo 1) Aquí hay dos manos.

(Moo 2) Si las manos existen, el mundo externo existe.

(Moo 3) El mundo externo existe.

De acuerdo con Wright, estos argumentos no resultan convincentes porque el sujeto que los formula no aprende la verdad de la conclusión a partir de la verdad de las premisas. El motivo es que su estructura presenta ciertos rasgos que impiden que la

²⁴¹ Debido al carácter trivial del defecto que presentan estos argumentos, en lo que sigue nos centraremos exclusivamente en el segundo y tercer tipo de fallo.

²⁴² El argumento GOL aparece en Wright (2000b), p. 142 y en (2002), p. 333 y el argumento VOTO en Wright (2002), p. 333. Este autor establece una vinculación entre estos argumentos y la prueba del mundo externo de Moore en (2003), pp. 336 y ss.

garantía aducida a favor de las premisas sirva para añadir garantía a la conclusión.²⁴³ Los rasgos relevantes de la estructura de los argumentos tipo (FTG 2) son dos. En primer lugar, las *evidencias* que pretenden garantizar las premisas son de carácter empírico (Wright 2000b, p. 147) y están *usadas de modo inferencial* (Wright 2003, p. 60).²⁴⁴ En segundo lugar, las evidencias que pretenden servir de justificación para garantizar las premisas sólo pueden cumplir este propósito si contamos con la verdad de *un conjunto colateral de informaciones al cual pertenece la conclusión del argumento*. Tomando como ejemplo el argumento GOL, Wright afirma que el griterío y los gestos de celebración, las evidencias empíricas que pretenden garantizar la primera premisa del argumento GOL (“se ha marcado un gol”), sólo garantizan dicha premisa de manera efectiva si aceptamos de antemano que se está celebrando un partido de fútbol, esto es, si asumimos previamente la verdad de la conclusión. Según Wright, cabría pensar en una situación en la que, más bien, se está representando teatralmente la disputa de un partido, en cuyo caso estarían disponibles para el sujeto que formula el argumento las mismas evidencias empíricas pero despojadas de su poder justificador.

En segundo lugar, hemos indicado que para Wright existe otra clase de argumentos no convincentes en los que se produce un fallo en la transmisión de la garantía que este autor no caracteriza en términos de petición de principio, esto es, el fallo que hemos denominado (FTG 3) y que es, precisamente, el problema que presenta el argumento de la consecuencia. De acuerdo con este autor, no es posible clasificar este argumento como (FTG 2) porque la conclusión no forma parte de la información

²⁴³ En relación a la caracterización del fallo que se produce en los argumentos tipo (FTG 2), existen algunas diferencias entre lo que Wright afirma en (2000b) y lo que sostiene después en (2003). De esta manera, en el primero de estos artículos se dice que sólo en aquellos argumentos que involucran premisas dependientes de información empírica podría cometerse petición de principio tipo (FTG 2). En este sentido, se afirma que “[l]o que es chocante sobre el caso de McKinsey [...] es que la evidencia no juega ningún papel para garantizar sus premisas [...] Puesto que no hay ninguna evidencia empírica en ningún caso, tampoco podría existir dependencia de información para la fuerza evidencial” (Wright 2000b, p. 147). Ahora bien, en (2003) este autor afirma que el factor diferenciador entre los razonamientos tipo (FTG 2) y los de tipo (FTG 3) es, más bien, que la dependencia de información colateral se presenta en caso de que la garantía tenga un carácter inferencial. Así, “[l]a plantilla de dependencia informativa se ajusta sólo a la transmisión de garantías inferenciales: garantías consistentes en la posesión de evidencia, e, que autoriza una inferencia falible una o más premisas de la implicación en cuestión” (Wright 2003, p. 60). De ahora en adelante, sólo tomaremos en consideración esta última caracterización que Wright ofrece sobre la garantía de los argumentos tipo (FTG 2).

²⁴⁴ Tomando como ejemplo el argumento GOL, podemos explicitar una inferencia que toma ciertas evidencias como premisas y la conclusión (Gol 1) en tanto que conclusión obtenida inferencialmente a partir de estas evidencias:

(Pre-gol 1) Escucho “gol”, veo gestos de celebración por parte de un conjunto de individuos vestidos de la misma manera y tristeza en los rostros de un segundo conjunto de individuos vestidos, todos ellos, de manera distinta a los primeros.

(Pre-gol 2) Si se dan las evidencias recogidas en (Pre-gol 1), se ha marcado un gol.

(Pre-gol 3) Se ha marcado un gol.

colateral que aportaría poder justificativo a las evidencias que se alegan a favor de las premisas y porque, además, estas evidencias no están siendo usadas de manera inferencial. Para entender en qué consiste este peculiar fallo en la transmisión de la garantía que es (FTG 3), vamos a atender a su formulación del argumento de la consecuencia que, de acuerdo con este autor, presenta esta clase de defecto. Para este fin enunciaremos la versión del mismo que nos propone este autor en (2000b, p. 144), y que, según nos indica, toma como referencia principal la propuesta de McKinsey:

ARGUMENTO CONSECUENCIA

(CoWri 1) Creo que el agua es húmeda.

(CoWri 2) Cualquiera que crea que el agua es húmeda (pertenece a una comunidad de hablantes que) ha tenido tales y cuales encuentros con el agua.

(CoWri 3) Yo (o mi comunidad) he (o hemos) tenido tales y cuales encuentros con el agua.

Una vez enunciado este argumento, hay que realizar dos precisiones. En primer lugar, es cuestionable que las comprensiones más usuales del externismo permitan establecer de manera no empírica la premisa (CoWri 2) y así lo reconoce el propio Wright.²⁴⁵ Según indicamos anteriormente, sólo será posible para aquellas versiones que consideren que la individuación externista de los contenidos de nuestros pensamientos convierte a estos últimos en dependientes de objeto.²⁴⁶ Sin embargo, Wright asume esta comprensión fuerte del externismo como el peor escenario posible para el externista con el fin de determinar si, en este caso, el argumento permitiría ganar conocimiento (*a priori*) de la conclusión (CoWri 3) a partir de sus premisas. La segunda precisión es la siguiente. De mantener una versión del externismo lo suficientemente fuerte como para implicar dependencia de objeto, ha de darse cabida a posibles “ilusiones de contenido”. La razón es que, de acuerdo con esta comprensión fuerte del externismo, un hablante que profiere una oración que contiene un término de género natural vacío como “fluido

²⁴⁵ Según hemos indicado, en nuestra opinión es bastante improbable que lo permita alguna de las versiones de las tesis externistas que estamos tomando en consideración en este trabajo. Si se trata del externismo natural, que afirma que nuestros conceptos de género natural reciben una individuación externa, habría que contar con una gran cantidad de información empírica para establecer (Co-Wri 2), información destinada a averiguar si el concepto expresado mediante el término “agua” es un concepto de género natural. Si se trata del externismo social, un individuo podría poseer el concepto de agua aun cuando ni él ni su comunidad hayan mantenido encuentros con esta sustancia, si bien se requeriría que él o un subconjunto de hablantes de su comunidad hubieran teorizado sobre el agua.

²⁴⁶ “El externismo sobre este tipo de conceptos [conceptos de género natural] puede involucrar todas o sólo algunas afirmaciones de entre esta variedad: por ejemplo, que el concepto está rígidamente asociado con su extensión actual, que se identifica con su extensión, que no puede existir sin una extensión (no vacía). Esta última es la que está vinculada con la segunda premisa de la paradoja de McKinsey” (Wright 2000b, p. 145).

calórico”, “éter”, “flogisto”, etc., no estaría expresando un genuino pensamiento. Ahora bien, en caso de que desconociera la vacuidad del término, este hablante podría creer, erróneamente, que expresa un pensamiento al decir “el éter llena el universo”. De ser así, nuestro individuo estaría sufriendo una “ilusión” al creer que su preferencia expresa un genuino contenido proposicional.

Visto esto, Wright justifica la clasificación del argumento CONSECUENCIA como un argumento de tipo (FTG 3) porque cualquiera que establezca de manera no-inferencial la verdad de la primera premisa se encuentra en un estado interno totalmente indistinguible por lo que a él respecta de otra situación (en la cual sufre una ilusión de contenido) que, de darse, haría que dicha premisa fuera falsa. En el caso del argumento CONSECUENCIA, la situación del hablante que, en efecto, tiene la creencia de que el agua es húmeda es totalmente indistinguible de aquella en la que el individuo vive en un mundo vacío, está sufriendo una ilusión de contenido y emplea el término “agua” aun cuando este concepto carece de extensión. Con el fin de aclarar al máximo la clase de defecto que instancian los argumentos de tipo (FTG 3), Wright examina dicha estructura del siguiente modo:

Sea un argumento que contiene A como una de sus premisas y B como su conclusión. En este argumento no se transmitirá la garantía de las premisas a la conclusión en el sentido (FTG 3) si:

- (i) A implica B.
- (ii) Mi garantía para A consiste en permanecer en un estado subjetivamente indistinguible de otro estado en el cual la alternativa relevante C sería verdadera.
- (iii) C es incompatible con A y
- (iv) C sería verdadera si B fuese falsa.

Puede apreciarse que el argumento CONSECUENCIA se adapta perfectamente a esta estructura. A sería la primera premisa “creo que el agua es húmeda”. De acuerdo con el externismo fuerte, esta premisa implicaría B, a saber, que mi comunidad o yo misma hemos tenido tales y cuales encuentros con el agua. Ahora bien, cuando afirmo A me encuentro en un estado subjetivamente indistinguible de aquél en el que afirmaré C, esto es, el estado en el que sufro una ilusión de contenido porque creo estar pensando que el agua es húmeda cuando en realidad no es así. Pero C, como hemos visto, excluye A y no permite concluir B. La cuestión es, en último término, que dado que ambos estados son subjetivamente indistinguibles resulta completamente arbitrario afirmar o

bien A (y concluir B) o bien C (y no poder concluir B). Para que el establecimiento de A (“creo que el agua es húmeda”) en el argumento de la consecuencia estuviese fundamentado, tendríamos que tener alguna justificación adicional para descartar C (“estoy sufriendo una ilusión de contenido”), lo cual supone, como se ha visto, descartar la negación de B (“ni mi comunidad ni yo misma hemos tenido tales y cuales encuentros con el agua”). Pero como descartar la negación de B no es más que afirmar B, resulta que para poder elegir A (y descartar C) con fundamento, tendríamos que contar de antemano con la verdad de B.

En resumidas cuentas, la tesis nuclear de Wright es que el argumento CONSECUENCIA constituye una instancia de cierto tipo de argumentos, los argumentos (FTG 3), que no permiten ganar conocimiento de su conclusión al no poder transmitirse la garantía por la ruta inferencial. Por este motivo las pretensiones anti-escépticas de dicho argumento quedarían frustradas de raíz y ningún agente que formule este argumento podrá convencer al escéptico.

Una vez expuestas las consideraciones de Wright en torno al argumento de la consecuencia, querríamos indicar cuál es nuestra posición acerca de sus tesis. En primer lugar, aunque pensamos que sus ideas sobre el escaso potencial anti-escéptico del argumento de la consecuencia son esencialmente correctas, alegaremos que no está claro que los argumentos GOL, VOTO u otros similares, catalogados por Wright como argumentos tipo (FTG 2), carezcan de carácter convincente para *cualquier* hablante *en cualquier contexto*. En nuestra opinión es posible pensar en situaciones, por lo demás muy usuales, en las que algunas instancias de estos argumentos podrían resultar perfectamente convincentes para el sujeto que las formula.²⁴⁷ Lo que trataremos de mostrar en lo siguiente es que estos argumentos pierden su potencial de convicción sólo ante cierta clase de dudas que, por lo demás, tienen un gran parecido con las que plantea el escéptico. En segundo lugar, mantendremos que es posible utilizar una noción de evidencia más amplia que la que Wright emplea en su caracterización de los argumentos

²⁴⁷ Este es un matiz de cierta importancia que, según pensamos, el propio Wright aceptaría. No obstante, su posición al respecto es un tanto ambigua. Por un lado, en (2000b), tras exponer el argumento GOL, este autor se pregunta si la garantía aducida a favor de las premisas puede transmitirse a la conclusión y procurar, así, un carácter convincente al argumento. Su respuesta es “no, al menos en el siguiente escenario” (Wright 2000b, p. 142). Y, a continuación, toma en consideración un contexto en el que un individuo se encuentra en las cercanías de un estudio de cine y donde la hipótesis de que se esté representando un partido de fútbol cobra plausibilidad. En este caso, diríamos, Wright parece aceptar la idea de que sólo en ciertos escenarios este argumento carece de poder convincente. Sin embargo, por otro lado, el tratamiento de Wright de estos argumentos indica que, en su opinión, GOL y VOTO, impiden, por su misma estructura, que la garantía aducida a favor de las premisas pueda transmitirse a la conclusión (Cf. Wright 2002, p. 333).

tipo (FGT 2) para elaborar una descripción unitaria de (las instancias de) los argumentos no convincentes de tipo (FGT 2) – que, según creemos, sólo carecen de carácter convincente en ciertas situaciones – y de instancias de los argumentos de tipo (FGT 3) como el argumento CONSECUENCIA.

Con respecto a lo primero, conviene comenzar describiendo una situación en la que podría tener lugar la enunciación del argumento VOTO por parte de un hablante. Se trata de un contexto en el que, en nuestra opinión, es bastante plausible pensar que dicho argumento convencería de la verdad de su conclusión al hablante que lo formula. Pensemos en la situación siguiente:

(Situación VOTO): Pepe entra por la puerta principal de la Facultad de Filosofía. Camina por el pasillo en dirección a la biblioteca y al final ve un numeroso grupo de personas. Con el fin de averiguar qué está pasando, se aproxima hacia la multitud y ve lo siguiente: su amigo Juan tiene en la mano una hoja donde puede leerse una lista de nombres encabezada por la palabra “candidatos”. A continuación escribe una “X” al lado de uno de esos nombres. Pepe formula para sí el siguiente argumento:

(Voto 1) Juan ha marcado una “X” en una papeleta electoral.

(Voto 2) Si Juan ha marcado una “X” en una papeleta electoral, Juan acaba de votar.

(Voto 3) Se están celebrando elecciones.

Por este argumento, Pepe infiere que se están celebrando elecciones.

Como puede apreciarse, en principio no parece haber nada inusual en un razonamiento de esta clase cuando se plantea en un proceso destinado a resolver la clase de dudas que surgieron en la mente de Pepe al ver a la multitud en el pasillo (“¿qué ocurre allí?”).²⁴⁸ De este modo, Pepe puede resolver sus dudas mediante la obtención de

²⁴⁸ Ante la objeción de que nuestra tesis depende únicamente de la fiabilidad de un solo ejemplo, podemos describir una situación semejante para contextualizar la formulación del argumento GOL:

(Situación GOL) Un individuo, Ángel, pasea por la calle muy cerca de un colegio durante un día no lectivo. Escucha cierto griterío y ve que hay gente en el interior del patio del colegio, visible desde la calle. Intrigado por este hecho, Ángel se acerca y ve a dos grupos de niños, bien diferenciados por el color de sus ropas. Los niños que pertenecen al primero de los grupos gritan “¡gol!” con expresión de júbilo, señalando un balón dentro de una portería, mientras que los del segundo grupo tienen una expresión triste. Ángel formula para sí el siguiente argumento:

(Gol 1) Se ha marcado un gol.

(Gol 2) Si se ha marcado un gol, se está jugando un partido.

(Gol 3) Por tanto, se está jugando un partido.

A partir de este argumento, Ángel concluye que los niños están jugando un partido de fútbol. Según creemos, esta contextualización del argumento GOL permite apreciar que el razonamiento de Ángel ha

la conclusión (Voto 3) a partir de las premisas del argumento relevante. No apreciamos en este contexto ningún síntoma de irracionalidad por parte del sujeto. Las percepciones de Pepe (ve una hoja de tales y cuales características, ve que Juan escribe una “X” al lado de un nombre señalado como “candidato”, etc.) le sirven para inferir que la hoja que su amigo Juan está rellenando es una papeleta electoral y, a partir de ello, que se está celebrando una votación. Por este motivo parece intuitivo pensar que el argumento VOTO, que, de acuerdo con Wright, carecía, por su propia estructura, de poder convincente, puede convencer al agente que lo formula en determinados contextos.

Conviene señalar, además, un segundo detalle, a saber, que la relación entre las percepciones de Pepe y la primera premisa del argumento GOL es una relación inferencial de *carácter abductivo*, esto es, la conclusión que obtenemos a partir de sus premisas (cuando éstas son verdaderas) es una conclusión *probable* y constituye, para quien formula el argumento, *la mejor explicación de los fenómenos descritos en la primera premisa*.²⁴⁹ Desde luego, podría ocurrir que Pepe hubiera tenido todas las percepciones descritas en dicha premisa y que, en realidad, se tratara sólo de una broma elaborada por Juan. Si ese fuera el caso, este agente no habría rellenado una papeleta electoral (ni se estaría llevando a cabo una votación), sino que sólo se estaría simulando tal cosa, aun cuando las acciones de Juan fuesen perceptivamente indistinguibles para Pepe de las que llevaría a cabo sobre una genuina papeleta electoral.

Llegados a este punto, aceptaremos que la lectura ofrecida de la situación de Pepe nos muestra que, al menos en algunas situaciones, un hablante puede darse por convencido formulando argumentos clasificados por Wright como (FTP 2). Para justificar esta tesis, examinemos, en primer lugar, qué peculiaridades presentan esta clase de situaciones. Según hemos visto, Pepe formula el argumento GOL para resolver un cierto tipo de duda, a saber, aquella que le incita a averiguar, en general, qué está sucediendo y a elaborar, en consecuencia, un argumento para obtener *de manera*

permitido resolver la duda que le surgió a este individuo al ver gente dentro de un colegio un día no lectivo (“¿qué estará pasando ahí?”).

²⁴⁹ Podemos esquematizar de la siguiente forma dicha inferencia abductiva:

(Pre-voto 1) Veo que Juan tiene en su mano una hoja, que en esa hoja se recogen una serie de nombres, que estos nombres están bajo el título “candidatos”, que Juan escribe una cruz al lado de uno de esos nombres.

(Pre-voto 2) Si Juan tiene en su mano una hoja de esas características y escribe una cruz al lado de uno de los nombres recogidos en la hoja de papel, entonces Juan está rellenando una papeleta electoral.

(Pre-voto 3) Juan está rellenando una papeleta electoral.

Como puede apreciarse, la conclusión (Pre-voto 3) ofrece la explicación más probable para los hechos enunciados en la primera premisa.

procesual una descripción (probablemente verdadera) de la situación que está presenciando. En un primer momento, como vimos, infiere la mejor explicación para sus percepciones y, después, utiliza la descripción inferida como primera premisa de un argumento cuya conclusión subsume el fenómeno o evento que está presenciando bajo una descripción más general que la descripción obtenida abductivamente.²⁵⁰

En segundo lugar, conviene establecer un contraste con otros contextos en los que, de manera igualmente intuitiva, algunas instancias de estos argumentos carecen de poder convincente, contextos que, en nuestra opinión, tienen un parentesco claro con los de aquellos hablantes que formulan argumentos clasificados por Wright como (FTG 3). Esto supone que aunque un agente puede quedar convencido de la conclusión de un argumento como GOL *en el contexto descrito* (i. e. para resolver esa clase de dudas de carácter general) ello no significa que Wright esté completamente equivocado, sino sólo que su diagnóstico de los argumentos de tipo (FTG 2) debe aceptarse *en relación a otros contextos* (i. e. para resolver otra clase de dudas). Ya hemos introducido la idea de que si se dieran ciertas condiciones externas, un tanto inusuales por lo demás, Pepe habría extraído una conclusión falsa del argumento abductivo elaborado para explicar sus percepciones. De este modo, podría ocurrir que las actividades de Juan, el agente que escribe una “X” en una hoja al lado de un nombre clasificado como “candidato”, formaran parte de una broma diseñada para llevar a Pepe a la creencia de que esa hoja es una papeleta electoral. Podríamos pensar en otro argumento de tipo (FTG 2) formulado en un contexto tal que haría que dicho argumento no resultara convincente:

²⁵⁰ Lo que queremos señalar con esto es que el vínculo entre la descripción ofrecida en el antecedente de la segunda premisa (obtenida abductivamente a partir de las percepciones) y la que figura en el consecuente (que es idéntica a la ofrecida en la conclusión de estos argumentos) es tal que, para que los fenómenos queden descritos con verdad bajo la segunda descripción, es condición suficiente que queden descritos con verdad bajo la primera. Esto es lo que ocurre, precisamente, cuando situamos una clase de fenómenos dentro de una clasificación de mayor generalidad que la primera. Así, si A es una clase que constituye un subconjunto de la clase B, todo objeto que pertenezca a A también pertenecerá a B. Si clasificamos ciertos hechos como “meter un gol”, también serán hechos clasificables como “disputar un partido de fútbol”. Para justificar esto, comprobemos que la función que tiene la segunda premisa de los argumentos VOTO y GOL – este último formulado en la nota anterior – es, en ambos casos, establecer una verdad puramente conceptual: si un grupo de gente rellena papeletas *electorales*, entonces se está llevando a cabo una votación (para *elegir* a alguien); si un equipo ha marcado un *gol*, entonces se está disputando un partido de fútbol (un deporte en el que se meten *goles*). De este modo, parece que lo establecido en las premisas (Voto 1) y (Gol 1) es algo *propio* de lo establecido en las conclusiones de estos argumentos, algo que sólo puede ocurrir en el contexto de lo descrito en (Voto 3) y (Gol 3). Si esto es así, una vez que Pepe y Ángel aceptan, a partir de sus percepciones, que lo más probable es o bien que Juan esté relleno una papeleta, o bien que los niños acaban de marcar un gol, parece inocuo concluir que Juan está participando en una votación y que los niños están jugando un partido de fútbol. Lo único que estos hablantes han hecho al concluir de esta manera es *clasificar* tanto la acción de Juan al rellenar la papeleta, como los gestos de alegría de los niños al marcar un gol, en el tipo de eventos en que estos hechos tienen su lugar.

(Situación ESTANCO): Pilar pasea por la calle y en un momento dado desea comprar tabaco. Situada delante de una tienda presidida por el logotipo propio de los estancos, Pilar se pregunta si en dicha tienda se venderá tabaco o sólo se trata de un local cuya apariencia externa es idéntica a la de los estancos usuales pero en la cual no se vende tabaco. Supongamos que el argumento que formula Pilar tiene la siguiente forma:

ARGUMENTO ESTANCO

(Est 1) Ese local de la esquina es un estanco.

(Est 2) Si es un estanco, se vende tabaco.

(Est 3) En ese local de la esquina se vende tabaco (no se trata de un estanco simulado).

Parece natural pensar que las percepciones de Pilar (como su visión del logotipo de la tienda) no servirían para establecer la premisa (Est 1), *dada su duda inicial*, a saber, si es una tienda donde se vende tabaco o si es una tienda donde no se vende tal cosa pero que resulta perceptivamente indistinguible de aquellas en las que sí se vende tabaco.²⁵¹

En el argumento ESTANCO nos encontramos con un hablante, Pilar que construye un argumento de tipo (FTG 2) para resolver una clase de dudas distinta de la que pretendía resolver Pepe. Según vimos, este último deseaba averiguar qué clase de fenómeno es aquél que está presenciando, algo que daría respuesta a preguntas muy generales como “¿qué pasa aquí?” o “¿qué está ocurriendo?”. Pero el propósito de Pilar es diferente. En particular, ella pretende *averiguar si lo que está presenciando es lo que parece*. Pilar sabe que la apariencia externa de la tienda ante la que está situada es típicamente la de un estanco y, por tanto, la de un establecimiento en el que se vende

²⁵¹ De nuevo, elaboramos un segundo argumento que demuestre que ESTANCO no es un caso excepcional. En este caso podríamos describir una segunda situación en la que dos individuos, Carlos y Javier, llegan a una plaza y se encuentran muy cansados. Carlos señala una hilera de piezas macizas de hormigón de unos 60 centímetros de altura y pregunta a su interlocutor si puede sentarse en una de esas piezas o si, más bien, se trata de obras de arte contemporáneo en las que está completamente prohibido sentarse. Javier, que nunca ha estado por esa zona ni ha visto a nadie sentarse en esas piezas, ofrece el siguiente argumento a su interlocutor para convencerle de que sí puede sentarse.

(Ban-1) Esa pieza de hormigón es un banco.

(Ban-2) Si es un banco, sirve para sentarte sobre él.

(Ban-3) Esa pieza de hormigón sirve para sentarte (no se trata de una obra de arte en la que está prohibido sentarse).

Ahora bien, cabe pensar que Carlos no se daría por convencido si supiera que su interlocutor tiene exactamente las mismas evidencias que él mismo para considerar que esa pieza es un banco genuino. De la misma manera que él no puede saber a partir sólo de sus percepciones si se trata de un banco o sólo de un “banco” artístico simulado, tampoco su interlocutor puede saberlo si cuenta con las mismas evidencias al respecto. En ese caso, es probable que Carlos reprochara a su interlocutor que ese argumento no ha resuelto *su duda*.

tabaco. Sin embargo, ella no quiere saber en general qué se vende en esa tienda (qué clase de tienda es), sino resolver la siguiente duda, a saber, si en ese establecimiento se vende tabaco (si es lo que parece) o si se trata más bien de una tienda que se parece mucho a un estanco pero donde no se vende tabaco (en cuyo caso no sería lo que parece ser).

En resumidas cuentas, constatamos que los argumentos de tipo (FTG 2), aunque resultan intuitivamente convincentes en los contextos de hablantes como Pepe, carecen de eficacia para agentes como Pilar, quien, ante un conjunto de percepciones, tiene una clase de duda muy determinada, a saber, *cuál de dos alternativas aparentemente indistinguibles es el caso*. Cabe decir, sin embargo, que, a menos que pueda alegar un motivo para creer que la situación externa es engañosa y, por tanto, dudar de las apariencias, Pilar sería un sujeto un tanto suspicaz, tanto que parece formar parte de los ejemplos típicos de los filósofos, *al igual que el escéptico*.

Ahora bien, llegados a este punto tenemos todos los elementos que nos permitirían defender que el argumento ESTANCO *en la situación descrita*, así como el argumento VOTO *en un contexto en el que Pepe dudara de si se está celebrando una votación genuina* (y no meramente simulada), parecen compartir algo muy importante con los argumentos en los que, de acuerdo con Wright, se produce un fallo en la transmisión de la garantía tipo (FTG 3). La razón es que el argumento CONSECUENCIA, clasificado por Wright como del tipo (FTG 3), está elaborado para resolver una duda similar a las dudas de Pilar y las de Pepe en ese caso. Para justificarlo, comprobemos que el argumento CONSECUENCIA utilizado con propósitos anti-escépticos busca expresamente resolver la clase de duda entre alternativas indistinguibles que pretendía resolver el argumento ESTANCO:

ARGUMENTO CONSECUENCIA

(CoWri 1) Creo que el agua es húmeda.

(CoWri 2) Cualquiera que crea que el agua es húmeda (pertenece a una comunidad de hablantes que) ha tenido tales y cuales encuentros con el agua.

(CoWri 3) Yo (o mi comunidad) he (hemos) tenido tales y cuales encuentros con el agua.

Dicho argumento, tal y como es usado por autores compatibilistas como Warfield y Sawyer, pretende extraer una conclusión anti-escéptica. Esto es, se usa con el propósito de responder a un hablante que plantea la siguiente duda: ¿estoy viviendo en un entorno que contiene agua o más bien soy víctima de un engaño masivo y en

realidad vivo en un mundo vacío? Este sujeto, al igual que Pilar, se pregunta *si las cosas son como parecen* o si más bien se encuentra en una situación alternativa donde todo le parece igual pero las cosas son radicalmente distintas. De la misma manera que el argumento ESTANCO no tiene eficacia para convencer de la verdad de su conclusión a un sujeto que plantea esta clase de dudas, el argumento CONSECUENCIA también es impotente en este sentido, dado el contexto (anti-escéptico) en el que se formula.

Llegados a este punto conviene recapitular las conclusiones obtenidas. En primer lugar, parece plausible que los argumentos de tipo (FTG 2) puedan convencer racionalmente a agentes que plantean dudas generales, del tipo mencionado, por lo que tales argumentos sólo carecerían de poder de convicción si el hablante que los formula o que demanda su formulación tuviera una duda en la forma de una alternativa como la descrita. En segundo lugar, el argumento CONSECUENCIA, de tipo (FTG 3), es un argumento que carece de poder de convicción por el mismo motivo que GOL y VOTO carecen de él en los contextos imaginados por Wright, a saber, porque tratan de resolver una duda entre alternativas indistinguibles.

Una vez expuestas estas primeras conclusiones, queremos argumentar a favor de la segunda tesis enunciada y, sirviéndonos parcialmente de las observaciones de García Suárez en (2000), defender que las razones que Wright ofrece para distinguir entre el fallo en la transmisión de la garantía de tipo (FTG 2), por un lado, y el fallo de tipo (FTG 3) que se presenta en el argumento CONSECUENCIA, por otro, son más débiles de lo que él pretende. Uniendo esta idea con las consideraciones precedentes sobre la dependencia contextual del carácter convincente de las instancias de los argumentos tipo (FTG2), pensamos que es posible elaborar una caracterización general del fallo en la transmisión de la garantía que dé razón de por qué los argumentos de tipo (FTG 2), debidamente contextualizados, carecen de poder de convicción por los mismos motivos que carece de dicho poder el uso anti-escéptico del argumento CONSECUENCIA – de tipo (FTG 3) –.

Para justificar esta tesis, en lo sucesivo sólo tomaremos en consideración los argumentos GOL, VOTO y demás ejemplos de argumentos tipo (FTG 2) en situaciones como las que indica Wright, esto es, en contextos en los que carecen de carácter convincente porque pretenden resolver dudas entre alternativas indistinguibles. En primer lugar, recordemos que, de acuerdo con Wright, un argumento instancia el tipo (FTG 2) si la conclusión del mismo forma parte del cuerpo de informaciones que servirían para otorgar un estatus justificador a las evidencias empleadas

inferencialmente para respaldar las premisas. Según este autor, el argumento CONSECUENCIA no podría clasificarse en el tipo (FTG 2) porque, en primer lugar, las garantías aducidas para las premisas no dependen de información colateral y, en segundo lugar, porque no quedan justificadas inferencialmente. Sin embargo, cabría pensar en una noción de evidencia más amplia que recogiera todo tipo de garantías, inferenciales o no. Si tomáramos esta decisión, podríamos dar un primer paso para demostrar que la distinción que realiza Wright entre el fallo en la transmisión de la garantía de tipo (FTG 2) y el fallo de tipo (FTG 3) que se da en el argumento de la consecuencia no es una distinción tan relevante como él defiende.

Para comprobarlo, vamos a elaborar una descripción de fallo en la transmisión de la garantía de un argumento más general que la de Wright y que retome las observaciones anteriores. A este respecto diremos que se produce un fallo en la transmisión de la garantía de un argumento si su conclusión pretende resolver una duda entre alternativas indistinguibles para el hablante (y para el oyente, si lo hubiera) pero es impotente para convencerle (o convencer al oyente) porque la verdad de dicha conclusión (que equivale a la elección entre una de las alternativas) debería presuponerse para considerar que las garantías (usadas inferencialmente o no) que se alegan a favor de las premisas permiten justificarlas de manera efectiva.²⁵²

Puede mostrarse que esta pauta se presenta tanto en los argumentos de tipo (FTG 2) – en los contextos indicados – como en los de tipo (FTG 3). En el primer caso, tanto si se está celebrando un partido de fútbol como si sólo se está representando, percibo *lo mismo por lo que a mí respecta*. Asimismo, tanto si soy un ser humano inserto en un entorno con agua como si soy un cerebro en una cubeta que nunca ha entrado en contacto con el agua, me encuentro en un estado *indistinguible por lo que a mí respecta* cuando “digo” para mis adentros “el agua es húmeda”. Mi estado interno no me permite decidir de manera no arbitraria si cuando uso el término “agua” estoy expresando una creencia sobre un líquido que fluye por ríos y lagos del mundo externo o si sólo estoy sufriendo una ilusión de contenido y no expreso tal creencia. Por este motivo, no podré convencer al escéptico de que debe escoger el primer miembro de la alternativa que plantea en su duda, a saber, que no vive en un mundo vacío. A partir de esto, podemos

²⁵² Según esto, no será convincente un argumento elaborado por un hablante que sea incapaz de otorgar un poder justificador a las evidencias aducidas a favor de las premisas a menos que se presuponga la verdad de la conclusión (que equivale a optar por una de las alternativas en duda), lo cual se requiere porque aunque la conclusión fuera falsa (y la alternativa descartada por él resultara ser verdadera), las evidencias aducidas a favor de las premisas serían completamente indistinguibles de las actuales por lo que a él respecta.

mantener que el uso de esta noción amplia de evidencia, que recoge tanto las evidencias empíricas como no empíricas en usos inferenciales o no, nos permite explicar de manera similar por qué los argumentos de tipo (FTG 2), como el argumento GOL en ese contexto, y los de tipo (FTG 3), son argumentos no convincentes.

El segundo paso para demostrar la similitud existente entre estos dos tipos de fallo en la transmisión de la garantía es hacer plausible que *también* en los argumentos de tipo (FTG 3), como CONSECUENCIA, la garantía de una de sus premisas depende de cierta clase de información colateral a la que pertenece la propia conclusión del argumento, lo cual constituye para Wright el segundo rasgo de los argumentos de tipo (FTG 2) (vid. García Suárez 2000, p. 168). En el caso del argumento GOL, debidamente contextualizado, se trata de la información que incluye, entre otras cosas, que se está celebrando un genuino partido de fútbol y, en el segundo caso, correspondiente al argumento CONSECUENCIA de tipo (FTG 3), se trata de un cuerpo de informaciones que incluye, entre otras cosas, que no me encuentro en una situación en la que sufro una ilusión de contenido. En cualquiera de esos dos casos, sólo si se contara con esa clase de información, las evidencias que sustentan la primera premisa de estos argumentos cobrarían fuerza justificativa y permitirían convencer a los hablantes que los formulan de la verdad de sus respectivas conclusiones.

Una vez hechas estas modificaciones, comprobamos que tanto los argumentos de tipo (FTG 2) en los contextos en los que no resultan convincentes, como el argumento CONSECUENCIA, son argumentos que carecen de poder de convicción por los mismos motivos, debido a que ambas clases de argumentos pretenden (i) convencer al agente resolviendo una duda entre alternativas indistinguibles a partir de (ii) evidencias (en sentido amplio) que son perfectamente compatibles con una explicación alternativa que, de ser cierta, haría falsa la conclusión del argumento, y fracasan porque (iii) sólo contando con un cuerpo de informaciones entre las que se incluye la propia conclusión del argumento podría otorgarse poder justificador a las premisas.²⁵³

²⁵³ Podemos reforzar nuestra tesis de que es posible ofrecer una explicación unitaria del carácter no convincente de los argumentos de los tipos (FTG 2) y (FTG 3) recorriendo el camino inverso. De este modo, si acabamos de mostrar que los argumentos de tipo (FTG 3) son muy similares a los de tipo (FTG 2) cuando estos últimos son usados para resolver cierta clase de dudas, también podemos explicar que los argumentos de tipo (FTG 2) en esta clase de contextos se adecuan a la esquematización del fallo en la transmisión de la garantía que presentan los argumentos de tipo (FTG 3). Recordemos qué forma tenía dicha esquematización:

- (i) A implica B.
- (ii) Mi garantía para A consiste en permanecer en un estado subjetivamente indistinguible de otro estado en el cual la alternativa relevante C sería verdadera.
- (iii) C es incompatible con A y

De admitir nuestras consideraciones, lo que Wright habría demostrado es que el argumento CONSECUENCIA y, más en concreto, el argumento de McKinsey, carece de poder de convicción frente al escéptico porque, *dada la clase de dudas que éste plantea ante las evidencias disponibles*, cualquiera que lo formule para convencer a este interlocutor de que vive en un mundo con agua (y no en un mundo vacío indistinguible del primero) fracasará en su intento de justificar la conclusión a partir de la verdad de las premisas.²⁵⁴

(iv) C sería verdadera si B fuese falsa.

Puede comprobarse que en los argumentos VOTO y GOL, que responden al tipo (FTG 2), también existen instancias de los elementos A, B y C, que, además, guardan entre sí las relaciones indicadas en el esquema. En el caso de VOTO, A sería “Juan está rellenando una papeleta electoral”, mientras que B sería “se está celebrando una votación”. Como puede verse, A implica B, lo cual queda recogido en la segunda premisa de este argumento. Por otro lado, podemos instanciar C como “Juan me está gastando una broma y finge rellenar un papel que parece una papeleta electoral pero no lo es”. De nuevo, C es incompatible con A: si Juan sólo finge rellenar un papel que parece una papeleta electoral pero no lo es, entonces Juan no está rellenando una genuina papeleta electoral. Igualmente, si C fuera verdadera, B no sería el caso, pues si lo que Juan rellena no es una papeleta electoral, entonces no se estaría celebrando una votación. En cuanto a la condición (ii) del esquema, también vemos que el argumento VOTO también la cumple. Tanto si Juan simula rellenar una papeleta falsa como si está rellenando de hecho una papeleta auténtica, las evidencias perceptivas de Pepe son exactamente las mismas. No obstante, Wright podría alegar que él usa la noción de “estado subjetivamente indistinguible” en su examen de los argumentos de tipo (FTG 3), mientras que nosotros acabamos de emplear la expresión “igualdad de evidencias perceptivas” al encajar el argumento VOTO en el punto (ii) del esquema. Sin embargo, consideramos que esta discrepancia no es demasiado importante. La razón de que Wright prefiera apelar a “estados subjetivamente indistinguibles” en el esquema recogido tiene que ver con el hecho de que las premisas de los argumentos de tipo (FTG 3) no quedan justificadas mediante garantías empíricas usadas inferencialmente, sino que, más bien, es ese estado interno el que jugaría el papel de la garantía. Ahora bien, hemos propuesto una noción amplia de evidencia para ver que es posible asimilar el fallo de tipo (FTG 2) al de tipo (FTG 3). Tanto si se trata de evidencias empíricas usadas inferencialmente como si apelamos a estados internos, la idea es la misma: por lo que al sujeto respecta, *todo le parece exactamente igual* (ya sea por sus evidencias empíricas o por del estado interno en el que se encuentra) tanto si es el caso que A (y por tanto B) como si es el caso que C (y por tanto no B).

²⁵⁴ No obstante, hay que decir que en algunos textos, como (2007) y (2002a), McKinsey señala que para que su paradoja sea efectiva no necesita contar con la transmisión de la garantía de las premisas a la conclusión, sino que sólo se requiere un cierto tipo de clausura epistémica que, según admiten Wright y Davies, sí se cumple en su argumento. En concreto, McKinsey denomina a este principio “principio de clausura de aprioricidad sobre implicación lógica” y enuncia su contenido de la manera siguiente: necesariamente, para cualquier persona x y cualesquiera proposiciones P y Q, si x puede saber *a priori* que P y P implica lógicamente Q, entonces x puede saber *a priori* que Q. No obstante, estamos de acuerdo con Brueckner (2005) en que la aplicación de este principio de clausura en vez del principio de transmisión de la garantía suscita exactamente los mismos problemas que planteaba inicialmente la formulación de su paradoja. Para comprobarlo, tengamos en cuenta que en estos textos McKinsey sostiene que los pensamientos que reciben una individuación externista tienen una forma lógica peculiar (relacional) tal que estos pensamientos implican lógicamente la existencia de aquellos hechos empíricos que los determinan. Esto supone que los pensamientos individuados en función de factores externos (como el pensamiento de que el agua es húmeda) tienen una forma lógica distinta de la de aquellos pensamientos que no reciben esta clase de individuación (como el pensamiento de que el éter llena el espacio). Si tenemos en cuenta que el nervio de la paradoja de McKinsey es que un individuo es capaz de saber *a priori* que existen los elementos empíricos que individualizan sus pensamientos, Brueckner señala que este individuo no podría saber tal cosa únicamente a partir del principio de clausura de aprioricidad. En su lugar, se requiere otro principio distinto que es el siguiente: necesariamente, para cualquier persona x y cualesquiera proposiciones P y Q, si x puede saber *a priori* que P y *sabe que* P implica lógicamente Q, entonces x puede saber *a priori* que Q. En este principio se exige que el individuo capaz de conocer *a priori* Q en tanto que consecuencia lógica de una proposición P ya conocida debe saber también, al

2.3.2 Consideraciones de Davies sobre el fallo en la transmisión de la garantía

Una vez concluido nuestro examen de las propuestas de Wright para desautorizar el empleo anti-escéptico del argumento de la consecuencia, vamos a retomar las ideas de Davies al respecto. Ya vimos que entre ambos autores existe una coincidencia fundamental, a saber, uno y otro autor mantienen que el argumento de la consecuencia es un argumento deductivamente válido, cuyas premisas son verdaderas, pero que es inservible para convencer de su conclusión a un agente, en tanto que el argumento presenta un fallo en la transmisión de la garantía de las premisas a la conclusión. Asimismo, sus propuestas presentan otro aspecto común, pues ambos toman en consideración una versión del argumento de la consecuencia relativamente similar, más parecida a la versión de McKinsey que a la de Boghossian y Brown, y que, en último término, opera sobre una concepción extremadamente fuerte del externismo. En particular, la versión del argumento que discute Davies es la siguiente:

(CoDav 1) Creo que el agua es húmeda.

(CoDav 2) Si estoy pensando que el agua es húmeda estoy (o he estado) inserto en un entorno que contiene porciones de agua.

(CoDav 3) Yo estoy (o he estado) inserto en un entorno que contiene porciones de agua.

Existe, eso sí, un primer punto de discrepancia entre la formulación de Davies y la de Wright, ya que este último reconocía la posibilidad de que la posesión del concepto de agua por parte del hablante que formula el argumento se debiera al hecho de que su comunidad (y no él mismo) haya tenido encuentros con el agua.²⁵⁵ Sin embargo, como vemos, la formulación de Davies no da cabida a dicha posibilidad, ya que ha de ser el propio sujeto que formula el argumento quien debe estar inserto en un entorno que contiene agua. Si, además, tenemos en cuenta que la existencia de porciones de este líquido en el entorno del hablante es una condición necesaria para que

contrario de lo que indica el principio de clausura de aprioricidad sobre implicación lógica, que Q es consecuencia lógica de P. El problema es que esto da lugar a la misma objeción que se planteaba originalmente frente al argumento de McKinsey de (1991), a saber, que un individuo es incapaz de saber *a priori* cuáles de entre sus pensamientos reciben una individuación externista. De ser esto cierto, un hablante no puede saber *a priori* si la forma lógica del pensamiento que está expresando mediante una preferencia determinada tiene la forma lógica del pensamiento de que el agua es húmeda más bien que la forma lógica del pensamiento de que el éter llena el universo.

²⁵⁵ Recordemos que la segunda premisa del argumento de la consecuencia propuesto por Wright era (CoWri 2): cualquiera que crea que el agua es húmeda (pertenece a una comunidad de hablantes que) ha tenido tales y cuales encuentros con el agua.

el hablante siquiera pueda pensar una proposición cuando asevera la oración “el agua es húmeda”,²⁵⁶ el resultado es que Davies también está pensando en una concepción del externismo que implica dependencia de objeto.²⁵⁷

A lo largo de este capítulo hemos indicado en varias ocasiones que la clase de teoría externista presupuesta en la formulación de la premisa (CoDav 2) resulta un tanto tosca y el propio Burge se negó a ver reconocidas sus propias tesis en dicha teoría.²⁵⁸ Sin embargo, asumiremos en lo siguiente la caracterización del externismo propuesta por Davies con el objeto de continuar el examen de la estrategia compatibilista de este autor, que articularemos en torno a tres tesis fundamentales. La primera es que existe una distinción entre la petición de principio y el fallo en la transmisión de la garantía. La segunda tesis es que ciertas instancias de argumentos cuyas premisas se establecen en base a garantías perceptivas pueden presentar fallos en la transmisión de dichas garantías, por lo que cabe ofrecer un Principio de Limitación en la Transmisión de la Garantía que se aplique en exclusiva a esa clase de argumentos.²⁵⁹ La tercera y última tesis es que el argumento de la consecuencia también presenta un fallo en la transmisión de la garantía pero, además, constituye una instancia de un problema específico, a saber, *el problema del conocimiento desde el sillón* (“from the armchair”). Con este peculiar calificativo, Davies se refiere al hecho de que este argumento presuntamente nos permitiría obtener conocimiento de cuestiones empíricas sin “levantarnos del sillón” y “salir” a la experiencia.²⁶⁰ El objetivo de Davies es construir un Segundo Principio de Limitación en la Transmisión de la Garantía que explique el tipo de fallo en la transmisión que presenta el argumento y que disuelva el puzzle planteado. Dicho de otra manera, si la garantía no puede transmitirse de las premisas a la conclusión, dicho

²⁵⁶ Cf. Davies (2000), p. 409 y (2003a), pp. 47 y 48.

²⁵⁷ Esto resulta aún más claro en textos como (2000, p. 406) donde este autor describe el externismo como una tesis sobre la dependencia lógica o conceptual de nuestros pensamientos respecto de ciertas condiciones externas, tal y como hacía McKinsey, quien, como vimos, mantiene una concepción del externismo entendido como una tesis sobre esta clase de dependencia lógica o conceptual. Vid. Davies (2003a), p. 45.

²⁵⁸ De hecho, el propio Burge afirmará al respecto que “a pesar de su carácter extremadamente esquemático, este principio [el principio de dependencia externista que es usado por Davies para formular el argumento de la consecuencia] – o cualquier instancia del mismo – es falso.” (Burge 2003b, p. 262).

²⁵⁹ Esta tesis de Davies debe entenderse de la siguiente manera. No es que usemos ciertos principios para limitar – en el sentido de *impedir* – la transmisión de la garantía de las premisas a la conclusión en cierta clase de argumentos (entre los que se encuentra el argumento de la consecuencia). Si así fuera la estrategia de Davies tendría un carácter puramente defensivo. Más bien, este autor pretende mostrar que la transmisión de la garantía en argumentos deductivos es un fenómeno restringido, limitado a un subconjunto de ellos. Con otras palabras, Davies busca enunciar con dicho principio en qué condiciones la validez lógica de un argumento no basta para que un individuo quede convencido de su conclusión aunque haya aceptado la verdad de las premisas.

²⁶⁰ Cf. Davies (2000), p. 387; (2003a), p. 105 y (2003b), p. 25.

argumento no permitiría en realidad aprender u obtener conocimiento de su conclusión (empírica) a partir de la aceptación de las premisas (establecidas desde el sillón).

Comenzando con la primera tesis de Davies, la distinción entre petición de principio y fallo en la transmisión de la garantía, cabe decir que, de acuerdo con este autor, la petición de principio es un fenómeno esencialmente *dialéctico*, un defecto que presentan ciertas instancias de argumentos ofrecidas para convencer a un oyente de la verdad de la conclusión. La idea de fallo en la transmisión de la garantía queda reservada para hablar de instancias de argumentos no convincentes que un oyente formula *para sí mismo*. La caracterización de Jackson de la noción de petición de principio, que Davies acepta plenamente, afirma que dicho fenómeno se produce cuando el oyente duda de la verdad de la conclusión y, al mismo tiempo, tiene creencias de fondo (“background beliefs”) cuya aceptación supone que la evidencia ofrecida para las premisas no tendría poder justificador (Jackson 1987, p. 107 y ss.). Davies señala que los argumentos GOL y VOTO pueden fracasar a la hora de convencer a un interlocutor en caso de que tenga las actitudes recogidas en la noción de petición de principio de Jackson porque, en ese caso, se cometería una petición de principio. Pero, además, también pueden fracasar a la hora de convencer al propio hablante que los formula cuando éste se encuentra en determinadas circunstancias (si tuviera ciertas creencias que desactivan el poder justificador de las evidencias alegadas a favor de las premisas), en cuyo caso un argumento como VOTO presentaría un fallo en la transmisión de la garantía.²⁶¹

Lo dicho hasta el momento nos permite establecer ciertas diferencias entre las tesis de Wright y las de Davies. Así, el primero consideraba que el fenómeno de la transmisión de la garantía incluye (pero no es coextensivo con) la comisión de una petición de principio. Mientras que los argumentos en los que la conclusión figura explícitamente en las premisas y los argumentos de tipo (FTG 2) son clasificados por Wright como argumentos que (además de no transmitir la garantía) cometen una

²⁶¹ Esto supone que no todas las instancias de argumentos como GOL y VOTO carecen de poder convincente (bien respecto a un interlocutor, bien respecto al propio individuo que los formula). De hecho, pensamos que las observaciones de Davies permiten pensar que en circunstancias ordinarias el argumento podría mover a convicción racional al menos a ciertos individuos, tal y como hemos defendido en las páginas precedentes. En los textos de este autor encontramos afirmaciones que confirman este extremo: “Veo lo que parece ser el progreso de un partido de la Liga Australiana de fútbol, y uno de los jugadores golpea el balón [y lo encaja dentro de los postes] [...] Grita la multitud, el árbitro hace un signo con sus manos [y pita]. Ordinariamente, esto contaría como una evidencia irresistible de que se ha marcado un gol. Así que yo tendría una garantía para creer la primera premisa del argumento [GOL] obviamente válido” (Davies 2000, p. 397). Nótese que hemos adaptado las observaciones de Davies para que encajen en el contexto de un partido de fútbol europeo, no americano.

petición de principio, los argumentos de tipo (FTG 3), como el argumento CONSECUENCIA, son argumentos donde se produce un fallo en la transmisión de la garantía que Wright no clasifica como una petición de principio. Por su parte, Davies considera que un mismo tipo de argumento (que en circunstancias ordinarias podría resultar convincente al menos para el sujeto que lo formula) puede dar lugar a instancias que incurren en una petición de principio si se formulan ante un destinatario *de ciertas características*, pero también podría dar lugar a otras instancias en las que se produce un fallo en la transmisión de la garantía siempre y cuando el hablante que los formula para sí mismo se encontrara en determinadas circunstancias que detallaremos en lo que sigue.

En nuestra opinión, el planteamiento de Davies nos parece más adecuado que el de Wright por dos razones. En primer lugar, la concepción dialéctica del fenómeno de petición de principio se acerca al sentido usual de dicho fenómeno, ya que resulta un tanto forzado decir que un hablante formula para sí mismo un argumento que comete una petición de principio. Por ello, preferimos reservar el problema del fallo en la transmisión de la garantía para explicar por qué un hablante formula para sí argumentos que carecen de poder de convicción. En segundo lugar, Davies hace mucho más hincapié que Wright en lo que respecta a la cuestión de qué clase de dudas y qué creencias relacionadas con esas dudas debe tener un sujeto para que un determinado argumento no le resulte convincente, un rasgo de sus propuestas que respaldamos enteramente.

Tras explicar la diferencia entre la petición de principio y el fallo en la transmisión de la garantía, vamos a atender a la segunda tesis de Davies, a saber, que es posible elaborar un Principio de Limitación de la Transmisión de la Garantía que explique por qué ciertas instancias de argumentos como GOL, VOTO o MOORE presentan un fallo en la transmisión de las garantías *empíricas* alegadas a favor de las premisas, de manera que convencen de la verdad de su conclusión al hablante que los formula.²⁶² Para ello, Davies ha propuesto distintas versiones de este Primer Principio,

²⁶² Dicho de otro modo, si ahora acabamos de explicar las peculiaridades del destinatario de estos argumentos cuando su formulación incurre en una petición de principio, ahora vamos a indicar cuáles son las características del individuo que formula para sí mismo estos argumentos y, sin embargo, no queda convencido de la verdad de su conclusión.

aunque sólo nos ocuparemos de las que ofrece en sus últimos escritos,²⁶³ ya que, según él mismo reconoce, las primeras versiones del mismo presentaban serias deficiencias.²⁶⁴

La versión genérica de este Primer Principio reza así:

VERSIÓN GENERALIZADA DEL PRIMER PRINCIPIO DE LIMITACIÓN DE LA GARANTÍA: La garantía epistémica no puede ser transmitida de las premisas de un argumento válido a su conclusión si, para una de las premisas, la garantía para dicha premisa cuenta como garantía sólo frente a ciertas asunciones de fondo y la aceptación (i) de esas asunciones y (ii) de las garantías para las otras premisas no puede ser combinada con la duda acerca de la verdad de la conclusión.

Explicemos este principio tomando en consideración una instancia del argumento VOTO. De acuerdo con Davies, un hablante que presente las actitudes descritas en este principio no podrá aprender la conclusión de dicho argumento a partir de sus premisas. En particular, las garantías perceptivas (“veo a Juan escribiendo una X en un papel al lado de un nombre clasificado como ‘candidato’”) que se alegan a favor de la premisa “Juan está rellenando una papeleta electoral” sólo pueden actuar como garantías si el hablante que formula este argumento cuenta ya con “ciertas asunciones de fondo”. El problema es que, en caso de que contara con tales asunciones, no podría, a su vez, dudar de la conclusión, es decir, dudar de que se estén celebrando elecciones.

Una vez expuesto el contenido de este Primer Principio en su versión genérica parece necesario comprobar si existen tensiones entre lo establecido por Davies y nuestras consideraciones precedentes. A este respecto, consideramos perfectamente legítimo que un sujeto formule un argumento como GOL o VOTO para obtener una descripción procesual de sus evidencias perceptivas a partir de una duda general como la que plantea la pregunta “¿qué está pasando?”. Si entendemos, como Davies y Wright proponen, que un argumento resulta convincente cuando el agente que lo formula aprende la verdad de la conclusión a partir de la verdad de las premisas, vemos que en estas situaciones las instancias del argumento VOTO y otros similares podrían resultar perfectamente convincentes.

Ahora bien, cuando Davies describe situaciones en las que cabe aplicar el Primer Principio de Limitación a sus argumentos plantea, al igual que Wright, *contextos* en los que las evidencias perceptivas estarían disponibles para el sujeto tanto si es el caso de lo

²⁶³ Nos referimos a Davies (2003a) y (2003b).

²⁶⁴ Vid. Davies (1998).

establecido en la conclusión como si no porque está teniendo lugar una situación alternativa a dicha conclusión pero compatible con tales evidencias y, además, el hablante que formula el argumento sabe esto o al menos lo sospecha. Así, Davies vuelve a hablar de una situación en la que un grupo de personas gritan “gol” y manifiestan gestos de celebración porque se está rodando una escena futbolística para una película en vez de disputarse un genuino partido de fútbol (Davies 2000, p. 397 y 2003b, p. 29). Si dicha situación fuera el caso y el hablante que formula el argumento GOL, ante tales evidencias, dudara de que se esté celebrando un partido, aceptaríamos el diagnóstico de Wright y Davies: dicho hablante no podría convencerse de la verdad de la conclusión (es decir, resolver su duda) mediante el argumento relevante.²⁶⁵

Una vez explicado el Primer Principio de Davies, vamos a tomar en consideración el Segundo Principio, que es el que nos permite explicar el fallo en la transmisión de la garantía que se produce en el argumento de la consecuencia. La diferencia entre el Primer Principio y el Segundo, recordemos, estriba en que la garantía ofrecida para las premisas de los argumentos a los que cabe aplicar el primero son de tipo perceptivo. El Segundo Principio de Davies es el siguiente.²⁶⁶

²⁶⁵ En este punto hemos de realizar una cierta precisión. En ocasiones Wright y Davies se expresan como si un individuo tuviera que asumir de antemano que no se encuentra en las inmediaciones de un plató de cine para que esas evidencias perceptivas cuenten como garantía de que se ha marcado un gol. Sin embargo, en situaciones ordinarias este requisito es demasiado fuerte. Parece demasiado pedir que un individuo que escucha un grito de “gol” y ve gestos de celebración deba creer de antemano cosas como que no está asistiendo a un ensayo de una película para que esas percepciones justifiquen su creencia de que se ha marcado un gol. Ahora bien, cabe interpretar las observaciones de estos dos autores de otra manera, esto es, como si estuvieran centrando su atención *únicamente* en la situación de un individuo que, ante esas percepciones, sigue dudando de que se esté disputando un partido de fútbol porque tiene alguna creencia (como “puede que sólo se trate de una representación”) que le impide emplear sus evidencias para justificar la premisa “se ha marcado un gol”. Pero, en nuestra opinión, lo que esto significa es que un individuo *ha de tener alguna razón* para no ofrecer la descripción “se ha marcado un gol” a la hora de explicar evidencias perceptivas como “escucho gritar ‘gol’”, “veo gestos de celebración por parte de un grupo de individuos vestidos con ropas del mismo color”, “veo un balón dentro de unos postes”, etc. Esto es así porque, usualmente, un individuo situado ante tales evidencias ofrecerá dicha descripción de manera prácticamente inmediata. Por estas razones, consideramos que la situación en la que Davies y Wright centran su examen resulta un tanto excepcional y se requiere, además, que el individuo tenga razones para pensarlo.

²⁶⁶ Este principio figura en Davies (2000), p.412; (2003a), p. 126 y (2003b), p. 45. Al igual que ocurría con el Primer Principio, hemos optado por recoger una versión del Segundo perteneciente a los últimos trabajos porque en ella se evitan los defectos de las primeras versiones.

VERSIÓN GENERALIZADA DEL SEGUNDO PRINCIPIO DE LIMITACIÓN DE LA GARANTÍA: La garantía epistémica no puede ser transmitida de las premisas de un argumento válido a su conclusión si, para una de las premisas, la aceptación de (i) la asunción de que hay tal proposición pensada por el hablante como premisa y (ii) las garantías para las otras premisas no puede ser combinada racionalmente con la duda acerca de la conclusión.

Veamos cómo se aplica este principio. El argumento de la consecuencia presenta este último tipo de fallo en la transmisión de la garantía porque quien lo enuncia (i) asume que hay una proposición expresada por la primera premisa (CoDav-1) – “creo que el agua es húmeda” – que constituye el contenido de la creencia del hablante, y acepta (ii) las garantías ofrecidas para la premisa (CoDav-2), que es la premisa sustentada en el externismo. El problema radica en que el hablante que admite los supuestos (i) y (ii) no puede dudar al mismo tiempo de la verdad de la conclusión. Dicho de otro modo, si un hablante cree (sin necesidad de justificación empírica) que está pensando que el agua es húmeda y, además, acepta el razonamiento de los externistas (establecido de manera no empírica), que hace depender dicha creencia de la existencia de agua en su entorno, entonces dicho sujeto no podría dudar en tal caso de que existe agua en su entorno.

El diagnóstico de Davies sobre el argumento de la consecuencia es el siguiente. Puesto que podemos aplicar este Segundo Principio a dicho argumento, éste no nos permitiría obtener conocimiento de hechos empíricos a partir de premisas justificadas de manera no empírica (“sin levantarnos del sillón”) por la sencilla razón de que forma parte de un tipo de argumentos que, aunque son lógicamente válidos, no permiten a quien los formula quedar convencido de la verdad de su conclusión.

Una vez expuesto el contenido fundamental de las tesis de Davies, querríamos realizar algunas observaciones al respecto. Es digno de destacar que este autor ha mostrado una preocupación más que evidente por una determinada objeción que podría hacerse a sus propuestas, a saber, que este Segundo Principio se ha enunciado puramente *ad hoc* con el fin de esquivar el problema incompatibilista del argumento de la consecuencia.²⁶⁷ En lo que sigue, elaboraremos una propuesta similar a la que presentamos en el caso de Wright, para obtener una explicación unificada del fallo en la transmisión de la garantía. Nuestro propósito es obtener un único principio de limitación

²⁶⁷ Quizá el texto más claro a este respecto es Davies (2000).

que quepa aplicar tanto a los argumentos GOL y VOTO²⁶⁸ como al argumento de la consecuencia. De lograrlo, consideramos que Davies podría resolver el problema del carácter *ad hoc* del Segundo Principio, puesto que no sería necesario establecer dos principios diferentes.

Para avanzar en el propósito mencionado atenderemos a una observación de Davies que, a nuestro juicio, tiene una considerable importancia, pero que, en nuestra opinión, no ha recibido suficiente atención. En particular, nos referimos a lo siguiente:

“Supongamos que tomamos el conocimiento *a priori* como un conocimiento que no descansa en la investigación empírica, pero que puede depender de *asunciones empíricas de fondo* (tales como la asunción de que “las condiciones para pensar un cierto pensamiento” se dan) [las cursivas son nuestras]. Esto hace mucho más plausible que las tesis de dependencia externista puedan ser conocidas *a priori* y mucho más que la autoridad de primera persona involucre conocimiento *a priori*. Pero aún así puede parecer absurdo que yo pueda saber *a priori* que estoy (o he estado) inserto en un entorno que contiene ejemplares de agua” (Davies 2003a, pp. 118-119).

Este fragmento de Davies se enmarca en un contexto en el que él está tratando de esclarecer qué noción de *a priori* se encuentra operativa en el argumento de la consecuencia. En particular, nos dice que las dos premisas de dicho argumento (las tesis externistas y el conocimiento en primer persona de un determinado pensamiento) son premisas que poseen un carácter *a priori* en el sentido de que puede establecerse su verdad sin necesidad de justificaciones empíricas, aunque *para ello se requiera contar con un conjunto de asunciones empíricas*. Esta precisión nos parece fundamental y encaja con lo que hemos señalado en las páginas precedentes, lo cual, en el fondo, no es más que lo que ha indicado el propio Burge en distintas ocasiones, a saber, que las tesis externistas no han sido establecidas *a priori* en el sentido de *a priori* que suponen los proponentes incompatibilistas del argumento de la consecuencia.²⁶⁹

²⁶⁸ Nos ceñimos a casos en los que se produce el fallo en los argumentos GOL y VOTO porque el sujeto que los formula no lo hace para extraer procesualmente una descripción general de sus percepciones en un entorno cuyo carácter fiable se da por supuesto.

²⁶⁹ “Por supuesto, sabemos que el agua es un género natural. Esto es conocimiento empírico de fondo. He usado tal conocimiento empírico de fondo al formar tesis sobre las condiciones bajo las cuales es posible tener el concepto de agua. He usado el conocimiento empírico de fondo de que existen compuestos y elementos. Puede decirse que incluso he asumido que el agua misma existe en algunos esbozos que he realizado acerca de las condiciones bajo las cuales es posible (o imposible) tener el concepto agua [...] [H]emos hecho uso de una gran cantidad de conocimiento empírico” (Burge 2003b, p. 263).

A fin de apreciar la importancia de este texto debemos reparar en que en el Segundo Principio no se hace mención alguna a la existencia de un conjunto de asunciones de fondo incompatibles con la duda sobre la conclusión, algo que sí ocurría en el caso del Primer Principio. Pero, al mismo tiempo, en el fragmento citado Davies sí reconoce la presencia de supuestos (empíricos, además) en el establecimiento del argumento de la consecuencia, al que cabe aplicar el Segundo Principio. Por este motivo creemos que el reconocimiento de la existencia de un cuerpo de asunciones empíricas de fondo operativas *también* en el argumento de la consecuencia puede proporcionarnos la clave para ofrecer una explicación unificada del fallo en la transmisión de la garantía que se produce en *todos* los argumentos examinados.

La versión del Principio de Limitación que proponemos, aplicable también al argumento de la consecuencia, sería la siguiente:

VERSIÓN UNIFICADA DE LA LIMITACIÓN DE LA GARANTÍA: La garantía epistémica no puede ser transmitida de un argumento válido a su conclusión si, para una de las premisas, la garantía (empírica o no) para dicha premisa sólo cuenta como garantía frente a ciertas asunciones de fondo y la aceptación (i) de esas asunciones y (ii) de las garantías (empíricas o no) para las otras premisas no puede ser racionalmente combinada con la duda acerca de la verdad de la conclusión.

Como puede observarse, esta Versión Unificada es más similar al Primer Principio propuesto por Davies que al Segundo. Lo único que se ha añadido es que las garantías para las premisas pueden ser de carácter empírico o no. Para justificar que esta Versión Unificada también explica el fallo en la transmisión de la garantía del argumento de la consecuencia debemos explicar las dos modificaciones realizadas sobre el Segundo Principio. La primera de ellas consiste en mencionar la existencia de un cuerpo de asunciones de fondo necesarias para el establecimiento y garantía de las premisas. En relación a esta primera modificación, estamos autorizados a realizarla desde el momento en que, según el fragmento que hemos citado en la página anterior, el propio Davies reconoce la necesidad de contar al respecto con ese cuerpo de asunciones de fondo también en el argumento de la consecuencia (2003a, pp. 118-119). Además, dichas asunciones desempeñan en este argumento exactamente el mismo papel que desempeñaban en los argumentos a los que cabe aplicar el Primer Principio, ya que en todos los casos *se trata de asunciones incompatibles racionalmente con la duda acerca de la verdad de la conclusión*. Esto puede comprobarse reparando en la asunción

existente en el argumento de la consecuencia que nos proporciona el propio Davies, a saber, que se dan las condiciones para poder tomar en consideración el pensamiento de que el agua es húmeda. Desde luego, teniendo en cuenta que para el externista dichas condiciones consisten en que haya agua en su entorno,²⁷⁰ es obvio que ningún agente podría combinar racionalmente la duda acerca de la verdad de la conclusión (“estoy inserto en un medio que contiene agua”) con la aceptación de (i) las tesis externistas y, especialmente, de (ii) la asunción empírica de fondo (“hay agua en mi entorno”) que haría que el establecimiento de la primera premisa estuviera realmente garantizado.

En cuanto a la segunda modificación efectuada, deberíamos argumentar por qué es legítimo sustituir el primer supuesto del Segundo Principio de Davies (aceptación de que hay una proposición pensada por el hablante como premisa) por el primer supuesto de nuestra Versión Unificada (aceptación de ciertas asunciones que permiten que las garantías para la primera premisa cuenten justamente como garantías). En este sentido, alegraríamos que si las garantías para establecer la primera premisa del argumento de la consecuencia contaran, en efecto, como garantías, esto supondría que un hablante estaría justificado a la hora de establecer la verdad de la primera premisa “creo que el agua es húmeda”. En caso de que “creo que el agua es húmeda” fuera una premisa verdadera, entonces el hablante que enuncia el argumento estará teniendo un pensamiento con el contenido proposicional de que el agua es húmeda o, lo que es igual, *habrá tal proposición pensada por el hablante*, justo lo que se enunciaba en el supuesto (i) del Segundo principio.

En resumidas cuentas, lo que hemos conseguido al plantear una Versión Unificada de la Limitación de la Garantía es eliminar la apariencia de principio *ad hoc* que tenía en Davies el Segundo Principio, usado para desactivar el argumento incompatibilista. De este modo, tanto los argumentos GOL, VOTO en las circunstancias señaladas, como el argumento MOORE y el argumento CONSECUENCIA, serán argumentos no convincentes porque presentan una misma clase de fallo en la transmisión de la garantía (sea o no de carácter empírico).²⁷¹

²⁷⁰ Como hemos indicado, de acuerdo con la versión de externismo sobre la que Davies formula su versión del argumento de la consecuencia.

²⁷¹ Conviene atender no obstante a una posible objeción a lo dicho hasta el momento. Así, en primer lugar, alguien podría preguntarse por qué estamos tan seguros de que los contextos en los que cabe aplicar este Principio son contextos en los que el hablante que plantea el argumento quiere resolver una duda de esta clase. En este sentido, ni Wright ni Davies mencionan una duda semejante y, sin embargo, nosotros hemos aceptado en lo fundamental sus conclusiones. Pues bien, esta pregunta puede responderse de la manera siguiente. En primer lugar, los ejemplos diseñados por estos autores plantean contextos en los que el hablante que formula los argumentos trata de resolver esa clase de dudas: ¿se está disputando un

3. Conclusiones

En este capítulo hemos tomado en consideración las distintas versiones del argumento de la consecuencia, un tipo de argumentación incompatibilista que pretende reducir al absurdo la conjunción del externismo y el acceso privilegiado con el objetivo de mostrar que, de admitir esa conjunción, tendremos que hacer frente a una consecuencia (en principio) altamente implausible, a saber, que podemos conocer *a priori* aquellos hechos empíricos que, de acuerdo con el externismo, determinan la identidad de los contenidos de nuestros pensamientos.

Las distintas versiones del argumento de la consecuencia han ido ganando en alcance y matices. Muchos autores coincidieron en que la primera versión del argumento ofrecida por McKinsey empleaba una noción de aprioricidad sumamente restrictiva y malinterpretaba las tesis externistas al describirlas como tesis acerca de la dependencia lógica o conceptual, más bien que metafísica, de nuestros pensamientos respecto de factores externos. En lo que se refiere a Boghossian, este autor ofreció su versión del argumento en el contexto de un estudio atento de las peculiaridades del externismo derivado del experimento mental de Putnam en torno al uso del término “agua”. Por último, Brown planteó una tercera versión del argumento de la consecuencia dirigida, esta vez, contra el externismo social de Burge, un tanto desatendido en las versiones anteriores del argumento.

Por otro lado, también hemos constatado que existe una idea común a la mayoría de los autores compatibilistas, a saber, que resulta extremadamente improbable que un

partido o sólo se está representando teatralmente un partido?, ¿se está celebrando una votación o sólo se está simulando tal cosa? Desde luego, existe una buena razón para que este tipo de dudas sea el cuestionamiento relevante en todos estos casos. Y es que lo más normal para un individuo que escucha “gol” y ve a un grupo de individuos con gestos de celebración es que infiera en ese mismo instante que se acaba de marcar un gol y que, por tanto, se está disputando un partido de fútbol. El tipo de razones que podrían obstaculizar esta ruta inferencial ordinaria serían aquellas que ponen en cuestión la fiabilidad del entorno. En este caso, si escucho gritar “gol” y veo a un grupo de individuos con gestos de celebración y, aun así, no describo la situación como la clase de situación que, de acuerdo con toda la evidencia disponible, parece estar dándose (“se ha marcado un gol”), sólo puede deberse a que dudo de que lo que está ocurriendo sea lo que parece, a saber, un partido de fútbol. Como hemos indicado, en este tipo de casos cabría aplicar la Versión Unificada de la Limitación de la Garantía. De este modo, si un individuo cree que su entorno es fiable y asume que no tiene por qué dudar de las apariencias, las garantías podrán justificar las premisas sin problemas. Sin embargo, de acuerdo con la Versión Unificada que estamos planteando, un hablante que se encuentre en esta situación no podría dudar racionalmente, al mismo tiempo, de la verdad de esa conclusión. Esto es, si un individuo escucha “gol”, ve gestos de celebración y asume que su entorno es fiable (que no se trata de una mera representación teatral), entonces no podría dudar de que se esté jugando un partido de fútbol, por lo que sería ocioso plantear el argumento GOL. En suma, si formulara el argumento en ese caso para convencerse de dicha conclusión, no estaría añadiendo a ésta ninguna garantía. El hablante, en última instancia, no aprendería nada, como señalan Wright y Davies.

sujeto pueda saber *a priori* cuáles de sus conceptos reciben una individuación externista. Este supuesto tan problemático era fundamental para la versión del argumento de la consecuencia propuesta por McKinsey. Boghossian, por su parte, trató de rebajar su exigencia alegando que basta con que un agente esté usando un término con *pretensiones* de expresar un concepto *atómico* de género natural para que pueda derivarse la conclusión del argumento. Pero, como indicamos, el requisito de que ese presunto concepto de género natural fuera un concepto *atómico* era una manera subrepticia de exigir que ese concepto fuese un concepto individuado de manera externista. Brown, por su parte, ofrece la versión del argumento de la consecuencia con un mayor alcance. La razón es que, de acuerdo con Burge, la gran mayoría de nuestros términos generales (y no sólo los términos de género natural) expresan conceptos que reciben una individuación externista. Además, dado que Burge reconoce que un hablante puede poseer un concepto cuya extensión es vacía pero cuyo uso estará sujeto a determinados criterios públicamente establecidos, Brown se apoya en esta observación para afirmar que, según el externismo social, un hablante que sabe *a priori* que posee un concepto puede deducir, asimismo, *a priori*, que existen otros hablantes responsables de establecer sus criterios de aplicación. Ahora bien, frente a esta afirmación, recogimos varios textos de Burge en los que se aclaraba que las tesis externistas son tesis que, en ningún caso, se han establecido *sin* contar con la existencia de un mundo externo o de otros hablantes. Por este motivo, al asumir de partida la existencia de otros hablantes que consensúan criterios de aplicación de nuestros conceptos, no resulta en absoluto paradójico concluir que, dado que poseemos conceptos y dado que esos criterios públicos de aplicación determinan su identidad, han de existir otros hablantes que los han establecido.

La segunda parte de este capítulo ha estado dedicada al examen de dos estrategias compatibilistas opuestas en gran medida. En primer lugar, analizamos las ideas de Warfield y Sawyer, dos autores que coinciden en aceptar la verdad y aprioricidad de las premisas y de la conclusión del argumento de la consecuencia pero que rechazan, al mismo tiempo, que dicha conclusión tenga un carácter paradójico. El origen de esta tesis puede rastrearse, como hemos indicado, en el uso que hace Putnam de su teoría de la referencia con propósitos anti-escépticos. Sin embargo, de la misma manera que cuestionamos la eficacia del argumento de Putnam, también hemos constatado que las pretensiones de Warfield y Sawyer son desmesuradas y la

combinación del externismo y el acceso privilegiado no nos sirve para refutar al escéptico.

En segundo lugar, tomamos en consideración las ideas de Wright y Davies, quienes, al igual que Warfield y Sawyer, aceptan las premisas y la conclusión del argumento de la consecuencia pero que, a diferencia de aquéllos, cuestionan seriamente la capacidad de este argumento para *convencer* de su conclusión a un escéptico. Aunque sus propuestas divergen en varios puntos, ambos coinciden en señalar que existe un fallo en la transmisión de la garantía de las premisas a la conclusión. Así, la clase de conclusión que se espera obtener de este argumento es tal que, de ser cuestionada por el escéptico, las evidencias disponibles no permitirían justificar las premisas. O, a la inversa, cualquiera que acepte que las evidencias disponibles constituyen una garantía suficiente para la verdad de las premisas es porque, en realidad, nunca ha puesto en duda la verdad de *esa* conclusión. En lo que a nosotros respecta, hicimos ciertas observaciones sobre las tesis de estos autores que nos invitan a poner más énfasis en el contexto de formulación de un argumento (en la clase de dudas que posee el agente que lo formula) para determinar en qué casos se produce un fallo en la transmisión de la garantía. Además, también señalamos que algunas de las distinciones de estos autores tenían un fundamento más débil de lo que podría parecer en un principio. Sin embargo, nuestro diagnóstico sobre la eficacia anti-escéptica del argumento de la consecuencia es esencialmente el mismo. Por una parte, si un escéptico planteara este argumento, no quedaría convencido de la conclusión porque, dada la clase de dudas que plantea este sujeto, las garantías alegadas a favor de las premisas quedarían en suspenso. Por otra parte, si un hablante no escéptico planteara este argumento, tampoco aprendería nada al extraer *esa* conclusión. En efecto, puesto que no duda de la conclusión, aceptaría sus premisas y las consideraría justificadas a partir de las evidencias disponibles, pero *sería ocioso* formular dicho argumento para convencerse de algo que nunca ha puesto en duda.

Apéndice: Las propuestas de McKinsey posteriores a (1991)

A lo largo de este cuarto capítulo hemos discutido el argumento presentado por McKinsey en (1991) contra las pretensiones compatibilistas. Según hemos señalado, su propuesta gozó de una gran influencia, no tanto por la fortaleza de su razonamiento en particular (sumamente cuestionable en algunos aspectos), como por el hecho de que inspiró argumentos más refinados a favor y en contra de la opción compatibilista. No obstante, hay que señalar que las pretensiones de ese trabajo fueron malinterpretadas de manera general. De esta forma, aunque la famosa paradoja únicamente buscaba poner de manifiesto que la conjunción del externismo y el acceso privilegiado da lugar a consecuencias implausibles, sin señalar cuál de estas dos tesis debe ser rechazada, la mayor parte de sus críticos y partidarios dio por sentado que el externismo sería la premisa más débil y que, en consecuencia, el argumento original de McKinsey tenía un carácter incompatible anti-externista.²⁷² Sin embargo, en los artículos de (2002), (2003) y (2007) McKinsey declaró que, en realidad, él era un firme partidario de la comprensión del externismo que intervenía en la generación de la paradoja y que aceptaba el resultado lógico de la misma, a saber, rechazar el acceso privilegiado a los propios pensamientos. El presente apéndice estará dedicado a examinar la argumentación de McKinsey en sus artículos posteriores a (1991) con el objetivo de arrojar la máxima claridad posible sobre las pretensiones y el alcance de la paradoja que propuso en ese trabajo.

En primer lugar, resulta conveniente especificar, por un lado, qué versión del externismo figura en la paradoja de (1991), puesto que ésa es la versión del externismo que McKinsey respalda, y, por otro, qué clase de acceso privilegiado es incompatible con él. En (2002) y (2007), este autor denomina “externismo semántico” (ES) a la primera de esas tesis y “principio de acceso al contenido” (PAC) a la segunda. El contenido de (ES) queda enunciado de la manera siguiente:

(ES): Algunos predicados estructurados *de dicto* de la forma “está pensando que P” expresan propiedades lógicamente amplias.

De acuerdo con McKinsey, una propiedad lógicamente amplia es aquella cuya atribución a un sujeto implica lógicamente (permite deducir *a priori*) la existencia de otros individuos distintos del sujeto de atribución. Como puede verse, este autor

²⁷² Brueckner (2004), p. 379 y (2007), p. 41.

mantiene en lo esencial la caracterización de dependencia externista (conceptual o lógica) que ofreció en (1991), según la cual si un hablante puede atribuirse con verdad un predicado como “pienso que el agua es húmeda” entonces sería posible para él deducir *a priori* que existe el agua en su entorno.

En sus artículos más recientes, McKinsey precisa que los predicados del tipo “piensa que P”, donde en P figuran nombres propios, indécicos o términos de género natural, son todos ellos predicados que expresan propiedades lógicamente amplias y que, por tanto, reciben una individuación externista “en sentido semántico”.²⁷³ La idea de este autor es que los predicados que expresan propiedades lógicamente amplias constituyen predicados relacionales con respecto al referente de los nombres propios, indécicos y términos de género natural. De ahora en adelante denominaremos “pensamientos_E” a aquellos pensamientos que atribuimos haciendo uso de predicados que expresan propiedades lógicamente amplias.

Por otro lado, el tipo de acceso privilegiado que interviene en la construcción de su paradoja es el *Principio de Acceso al Contenido*:

(PAC): Es necesariamente verdadero que si una persona x está pensando que p, entonces x puede en principio llegar a saber *a priori* que está pensando que p.²⁷⁴

Según pretende mostrar McKinsey mediante su famosa paradoja, (PAC) es incompatible con el externismo semántico porque la conjunción de ambas tesis permite deducir puramente *a priori* un hecho empírico, a saber, la existencia de aquellos individuos requeridos para la atribución de un pensamiento_E. De este modo, si un hablante pudiera conocer *a priori*, en virtud de (PAC), que está pensando un pensamiento_E, entonces podría atribuirse *a priori* un predicado que expresa una propiedad lógicamente amplia, esto es, una propiedad que implica *a priori* la existencia de otros individuos distintos de él mismo, es decir, del sujeto de atribución. Pero como es absurdo que un individuo pueda deducir *a priori* esta conclusión empírica, entonces debemos renunciar o bien a (PAC), o bien a (ES). Como McKinsey considera que el externismo semántico enunciado en (ES) es una tesis bien respaldada por “hechos

²⁷³ Cf. McKinsey (2002), p. 214 y (2007), p. 56.

²⁷⁴ Repárese en que McKinsey denomina “Principio de Acceso al Contenido” a un tipo de acceso privilegiado que en su enunciación no alude al conocimiento *a priori* del contenido del pensamiento p, sino al conocimiento *a priori* de p mismo. Esta puntualización es relevante para las consideraciones posteriores.

semánticos sobre nombres propios y términos indéxicos”, la tesis responsable de la inconsistencia sólo puede ser (PAC).²⁷⁵

Ahora bien, este autor se muestra reacio a asimilar la variante del externismo que él respalda a la comprensión usual del mismo. En este sentido, McKinsey se ve a sí mismo como un partidario del externismo en semántica, recogido en la tesis (ES), pero que, al mismo tiempo, rechaza la extrapolación de las tesis externistas al ámbito de la psicología. Aplicar el externismo a la individuación del contenido de nuestros pensamientos es respaldar una segunda variante del externismo, el “externismo metafísico” o (EM):

(EM) En algunos casos, una persona está pensando que *p*, el contenido que *p* es lógicamente amplio y el pensamiento de la persona está individuado por la propiedad de ser un pensamiento que tiene el contenido de que *p*.²⁷⁶

Según McKinsey, el externismo metafísico no sólo es incompatible con (PAC), bajo pena de permitirnos conocer *a priori* la existencia de los objetos externos requeridos para individuar el contenido del pensamiento de esa persona, sino que *también* entra en tensión con otra forma de entender el acceso privilegiado. Denominemos *Principio de Acceso Privilegiado a las Propiedades Individuadoras*, (PAI), a esta segunda clase de privilegio epistémico de primera persona:

(PAI) Es necesariamente verdadero que si el pensamiento de una persona está individuado por una propiedad dada *F*, esa persona puede en principio llegar a saber *a priori* que tiene ese pensamiento con la propiedad *F*.

La tesis (PAI) enuncia que una persona puede conocer *a priori* todas las propiedades que, como *F*, están involucradas en la individuación de uno de sus pensamientos. De acuerdo con McKinsey, la comprensión del externismo enunciada en (EM) es incompatible con (PAI) por un motivo muy similar al que hace incompatibles

²⁷⁵ Cf. (2002), p. 205 y (2007), p. 56.

²⁷⁶ Esta enunciación del externismo metafísico es la que figura en McKinsey (2002), p. 213. Hay que decir, no obstante, que en (2007), McKinsey denomina “externismo lógico” a (ME), puesto que caracteriza el contenido de *p* como *lógicamente* amplio, y lo distingue de una hipotética concepción metafísica del externismo que considera el contenido de *p* como *metafísicamente* amplio. Sin embargo, a continuación afirma que entender el externismo como una tesis metafísica carece de sentido: “no veo ninguna manera de que cobre sentido la afirmación de que los contenidos de algunos pensamientos dependan metafísica pero no lógicamente para su existencia de objetos contingentes o sustancias externas al agente” (McKinsey 2002, p. 60). De esta manera podemos decir que, para McKinsey, si la individuación externista de nuestros pensamientos es una tesis consistente, entonces debe ser entendida como una tesis que considera que nuestros contenidos mentales son lógicamente amplios, en el sentido de que implican *a priori* la existencia de individuos distintos de aquellos a los que se les atribuyen pensamientos con tales contenidos. Por esta razón, sólo tomaremos en cuenta la enunciación que McKinsey propone en (2002) de las tesis externistas “metafísicas” que él combate.

las tesis (ES) y (PAC). De este modo, si sostenemos (PAI) y afirmamos, con ello, que un hablante puede conocer *a priori* las propiedades individuadoras de uno de sus pensamientos_E (como que es un pensamiento con el contenido de que P), entonces este hablante es capaz de deducir *a priori* la existencia de otros individuos distintos de sí mismo. Como esta capacidad parece implausible, dice McKinsey, entonces sólo nos queda negar que un individuo tenga acceso *a priori* a la posesión de contenidos lógicamente amplios.

Visto esto, enunciaremos la primera de las dos tesis centrales que mantiene McKinsey tanto en (2002) como en (2007), a saber, que es posible declararse externista en el sentido “semántico”, según (ES), y, al mismo tiempo, rechazar la aplicación del externismo a la individuación de los pensamientos establecida en (EM). Para este autor aceptar (ES) sólo nos compromete con (EM) si aceptamos una tesis adicional, la “individuación por el contenido”:

(IC): Es necesariamente verdadero que si una persona está pensando que p, entonces su pensamiento está individuado por la propiedad de ser un pensamiento que tiene el contenido de que p.

La razón es que el externismo metafísico o (EM), que es la variante de externismo que McKinsey rechaza, es, en realidad, la conjunción de dos tesis, a saber, que (i) atribuir un pensamiento_E es atribuir una propiedad lógicamente amplia – algo aceptado por cualquiera que acepte la tesis (ES) – y que (ii) la propiedad de ser un pensamiento con un determinado contenido interviene en la individuación de dicho pensamiento, justo lo que enuncia la tesis (IC). Si renunciamos a (IC), podemos aceptar (ES) sin adherirnos a (EM), aunque no ocurre lo contrario: si asumimos (EM) debemos adherirnos a (ES).

La segunda tesis fuerte de McKinsey tanto en (2002) como en (2007) es, como hemos adelantado, que no sólo resulta posible desligar el externismo semántico del metafísico, sino que debemos aceptar sólo la primera de estas tesis y rechazar la segunda. El argumento que aporta el autor para tomar esta decisión es que la tesis (ES) cuenta con fuertes evidencias a su favor, tan firmes que si se demuestra su incompatibilidad con cierto tipo de acceso *a priori* a los propios pensamientos, (PAC), sólo nos queda rechazar esta clase de acceso *a priori*. Al mismo tiempo, la tesis (EM) no sólo resulta incompatible con el acceso *a priori* recogido en (PAC) – debido a que (EM) implica (ES) –, sino también con el tipo de acceso *a priori* enunciado en (PAI). McKinsey nos propone, pues, adherirnos a (ES) y rechazar (EM), puesto que de

respaldar (además) esta última tesis, estaríamos atentando contra el acceso *a priori* a los propios pensamientos en todas sus variantes y no sólo en (PAC).

Una vez recogido el esquema argumentativo de McKinsey en sus obras posteriores a (1991) es conveniente realizar algunas observaciones al respecto. En primer lugar, McKinsey no parece haber tomado en consideración una de las principales alegaciones compatibilistas a los supuestos de su paradoja, ya que continúa ofreciendo una caracterización del privilegio epistémico de primera persona en términos de aprioricidad y, lo que es más grave, sigue empleando una noción de conocimiento *a priori* extremadamente restrictiva al ofrecer, de nuevo, una descripción del conocimiento *a priori* muy similar a la que figuraba en (1991): “por conocimiento ‘*a priori*’ quiero decir conocimiento que es obtenido ‘sólo con pensar’ y no sobre la base de investigación empírica u observación perceptual” (McKinsey 2007, p. 54).

En segundo lugar, resulta un tanto difícil vincular la propia posición de McKinsey en tanto que externista con la de otros autores que se consideran a sí mismos partidarios del externismo. Sin embargo, podríamos abstenernos de cuestionar, por mor del argumento, qué relaciones existen exactamente entre la posición de otros autores y las variantes del externismo que discute McKinsey y preguntarnos por la viabilidad que cabe asignar a su posición particular. En nuestra opinión, el aspecto más débil de su argumentación es el supuesto de que el acceso privilegiado a los contenidos de nuestros pensamientos_E en la forma (PAC) es independiente del acceso privilegiado a las propiedades que individúan tales pensamientos (PAI).

La clave para desligar una clase de acceso *a priori* de la otra reside, como vimos, en el rechazo de McKinsey a la tesis de que nuestros pensamientos_E quedan individuados por (la propiedad de) tener el contenido que tienen, tal y como enuncia (IC). Ahora bien, negar que nuestros pensamientos estén individuados por su contenido es oponerse a una de las premisas que subyacen al debate entre externismo/internismo. Desde luego, si McKinsey tuviera razón y consiguiera demostrar que este principio ha viciado la discusión, no sería razonable respaldarlo por mero conservadurismo. Sin embargo, estimamos que el motivo para asumir la tesis (IC) es que goza de una plausibilidad muy intuitiva, en el sentido de que nos permite conectar el debate filosófico sobre la identidad de los contenidos de nuestros pensamientos con nuestros criterios ordinarios de atribución de los mismos. Así, en nuestras prácticas habituales de atribuciones psicológicas coincidimos en que Juan y María tienen pensamientos con el mismo contenido cuando dicen “el platino es más valioso que el oro” si pensamos que

ambos sujetos creen *lo mismo*, esto es, cuando uno y otro mantienen creencias con el mismo contenido. Negar (IC) supone establecer que la identidad/diferencia de contenidos no determina la identidad/diferencia de nuestros pensamientos y, como vemos, esto conllevaría alejarse de una de las pautas que sustentan nuestras prácticas ordinarias de atribución de pensamientos.

Sin embargo, no es éste el principal motivo por el que consideramos inapropiada la propuesta de McKinsey. A nuestro modo de ver, no tiene demasiado sentido afirmar que sabemos *a priori* qué propiedades individuán nuestros pensamientos, como reza (PAI) y, al mismo tiempo, mantener que no podemos saber *a priori* si mantenemos o no un determinado pensamiento, esto es, negar (PAC). Para justificarlo, tengamos en cuenta que la conjunción de ambas afirmaciones supone que un sujeto puede saber *a priori* qué hace que los pensamientos que mantiene sean los pensamientos_E que son pero, al mismo tiempo, no puede saber *a priori* que los mantiene. En efecto, desde el momento en que McKinsey se declara partidario de (ES), la posesión de un pensamiento_E no es algo que un hablante pueda atribuirse *a priori*, ya que esta clase de pensamientos implican lógicamente la existencia de otros individuos distintos, un hecho no cognoscible *a priori*. La pregunta que debemos hacernos es si resulta viable atribuir a un hablante la capacidad de saber *a priori* qué pensamientos_E tiene (al tener acceso *a priori* a sus propiedades individuadoras) pero negar que pueda saber *a priori* si los tiene. En este sentido, sólo se nos ocurren dos alternativas para conjugar estas dos tesis, ambas igualmente perjudiciales para los propósitos de McKinsey.

En primer lugar, podría parecer que para que uno sea capaz de instanciar *a priori* (PAI) en relación con un pensamiento_E dado y conocer, con ello, las propiedades individuadoras de *ese* pensamiento_E, se necesitaría saber que, en efecto, uno posee dicho pensamiento_E. Sin embargo, esto supondría que la instanciación dejaría de producirse puramente *a priori* y, por tanto, implicaría rechazar el principio del acceso *a priori* a las propiedades individuadoras o (PAI). La razón es que saberse poseedor de ese pensamiento_E determinado es algo que sólo puede conocerse *a posteriori*. Por ello, un hablante sólo podría saber qué propiedades individuán un pensamiento_E a partir de indagaciones empíricas, con lo cual violaríamos uno de los requisitos de la noción de aprioricidad que maneja McKinsey, a saber, prescindir por completo de investigaciones empíricas en el conocimiento de las propiedades individuadoras de los pensamientos que nos auto-atribuimos.

Por otro lado, cabe una segunda vía para intentar conjugar (PAI) y la negación de (PAC), a saber, mantener que yo no conozco *a priori* las propiedades que individúan el pensamiento de que P, sino que más bien conozco *a priori* qué propiedades individuarían ese pensamiento en caso de que lo tuviera. Es decir, que para conjugar estas dos tesis sólo se requiere que un hablante sepa *a priori* qué determinaría que mantuviese el pensamiento de que P o, lo que es igual, qué pensaría si estuviera pensando que P. No obstante, estimamos que una solución de este cariz tendría un alcance muy corto. A este respecto, intentaremos reconstruir el monólogo interior que un lector de McKinsey podría establecer reflexionando *a priori*. Dicho monólogo discurriría así:

“Según entiendo a este autor, puedo conocer *a priori* las propiedades individuadoras de mis pensamientos_E pero, al mismo tiempo, no puedo saber *a priori* que estoy manteniendo un pensamiento de esa clase. Ahora bien, si McKinsey entiende ‘individuar’ como yo lo entiendo, entonces está diciendo que puedo conocer *a priori* las propiedades que fijarían la identidad de mis pensamientos_E, esto es, las propiedades que determinarían qué pensamientos_E son los míos, en caso de que los tuviera. De este modo, en virtud de (PAI), puedo saber *a priori* qué pensamiento_E pensaría – en caso de que lo pensara – al emitir la preferencia ‘el oro es valioso’, a saber, *el pensamiento de que el oro es valioso*. Pero esto me llena de confusión porque, según parece, para poder reflexionar de esta manera necesito pensar ese pensamiento”

Lo que este monólogo interior ilustra es la dificultad que existe para mantener que soy capaz de conocer *a priori* las propiedades que individúan mis pensamientos, como enuncia (PAI), pero que, por todo lo que sé *a priori*, sólo puedo considerar de manera hipotética que tengo algún pensamiento individuado por estas propiedades. En nuestra opinión, no podemos desligar tan fácilmente una facultad de otra. La razón es que aun cuando la forma de acceso privilegiado que McKinsey respalda, (PAI), sólo autoriza el conocimiento *a priori* de formulaciones condicionales como “sé qué pensaría si estuviera pensando *que P*”, para poder pensar *a priori* esa clase de formulaciones se requiere que el sujeto esté tomando en consideración *a priori* ese mismo pensamiento_E de que P y, por tanto, que sea capaz de atribuirse *a priori* a sí mismo dicho pensamiento_E.

Nuestra objeción, en resumen, descansa en el hecho de que si McKinsey quiere independizar (PAI) de (PAC) y sostener la primera tesis sin comprometerse con la

segunda, sólo podrá hacerlo al precio de vaciar de significado el término “propiedades individuadoras”. Tal y como suele entenderse esa expresión, conocer las propiedades que individúan uno de nuestros pensamientos significa que tenemos acceso a los factores que hacen que uno de mis pensamientos sea *ese* pensamiento (el pensamiento de que P) y no cualquier otro. Dado mi conocimiento *a priori* de un conjunto de propiedades individuadoras de esta clase, me encuentro ante una alternativa cada uno de cuyos miembros son igualmente perjudiciales para los propósitos de este autor, a saber, o bien esas propiedades individúan uno de mis pensamientos y entonces tener acceso reflexivo *a priori* a esas propiedades como determinantes de *ese* pensamiento no puede desligarse de saber *a priori* que lo estoy pensando; o bien esas propiedades no son, en realidad, genuinas propiedades individuadoras.

Capítulo 5: Hacia una concepción viable de la individuación externista de los contenidos mentales

En los capítulos precedentes hemos tomado en consideración los distintos argumentos formulados en torno a la posibilidad de compatibilizar la individuación externista de los contenidos mentales con el acceso privilegiado a los mismos. A partir de una definición de mínimos del externismo psicológico, obtenida en un primer capítulo dedicado al examen de los fundamentos semánticos de esta tesis, hemos articulado el debate en torno a tres argumentos de progresiva complejidad, a saber, la incompatibilidad *prima facie* entre externismo y acceso privilegiado, los casos de traslado inadvertido y, por último, el argumento de la consecuencia.

En el presente capítulo nos proponemos abordar las siguientes cuestiones. En primer lugar, dedicaremos una sección a analizar las diversas tesis que se agrupan bajo la denominación de “externismo psicológico”. Al comienzo de nuestro trabajo ofrecimos una caracterización del externismo lo suficientemente amplia como para permitirnos recoger qué es lo común a las diversas teorías externistas sobre la individuación de los contenidos mentales, a saber, que toman en consideración factores externos a los hablantes, de carácter natural o social, para *determinar la identidad* de dichos contenidos. No obstante, también hemos podido apreciar que esta comprensión puede instanciarse de diversas maneras y que, además, algunos autores que han participado en la polémica compatibilista han denominado “externismo psicológico” a tesis que no tienen que ver estrictamente con la individuación de los contenidos mentales. Por este motivo, vamos a comenzar este quinto capítulo ofreciendo un esquema y un análisis de las opciones teóricas que, presentándose bajo el rótulo de “externismo psicológico”, se han posicionado sobre diversas cuestiones relacionadas con los estados y procesos mentales.

En segundo lugar, nos centraremos en el examen de aquellas tesis que, de acuerdo con el esquema que ofrecemos, se encuadran exclusivamente dentro del externismo con respecto a la individuación de nuestros contenidos mentales, la tesis cuya compatibilidad con el acceso privilegiado se cuestiona. Reservaremos, pues, una segunda sección a analizar la clasificación usual de las distintas variedades de esta clase de externismo, así como a explicitar aquellos fenómenos relacionados con el lenguaje y la atribución de pensamientos que pretenden servir de apoyo a las tesis externistas. El fin que perseguimos es explicar por qué esta tipología tradicional de las concepciones

del externismo sobre la individuación no resulta del todo apropiada. Según argumentaremos, no es tan sencillo conjugar la influencia de los factores de tipo social, por un lado, y la influencia de los factores de tipo natural, por otro, en la determinación de los contenidos psicológicos. Para ello, nos serviremos de las ideas de algunos autores que han señalado que reconocer la influencia de un tipo u otro de factores requiere apoyarse en intuiciones muy distintas y que, en ocasiones, arrojan resultados contrarios.

A partir de estas observaciones, en la tercera sección elaboraremos una clasificación alternativa de las variantes de externismo. Nuestra preocupación no será tanto determinar qué autores han mantenido alguna de estas concepciones, como el hecho de que en esta nueva tipología las tesis distinguidas evitan, al menos en un principio, los conflictos entre la influencia de los factores naturales y la influencia de los factores sociales en la determinación de los contenidos mentales. Tras ofrecer esta nueva clasificación nos proponemos argumentar contra aquellas concepciones externistas que reservan un poder individuador a los factores naturales no mediado por los usos sociales de nuestros términos. En este sentido, pensamos que la teoría de la referencia de los términos de género natural que sirve de sustento a la mayor parte de las concepciones del externismo psicológico – en especial, a la variedad denominada “externismo estrictamente natural” – resulta extremadamente cuestionable y, además, es la responsable de los conflictos más difícilmente resolubles entre el externismo y el acceso privilegiado. Por este motivo, daremos cuenta de las dificultades específicas que menoscaban la viabilidad del “externismo estrictamente natural”, retomando para ello el experimento mental de la Tierra Gemela en torno al uso del término “agua” y la teoría de los géneros naturales y de la referencia de los términos de género natural que constituye el apoyo teórico de esta clase de externismo. La razón de nuestro interés reside en que esta tesis no sólo sustenta el externismo estrictamente natural, sino también otras variedades que distinguiremos a lo largo de este capítulo, como son el externismo de distribución y el externismo social esencialista. Si nuestras consideraciones son correctas, todas estas concepciones tendrán que hacer frente a una serie de problemas que no afectan al externismo social no-esencialista, la tesis por la cual abogamos.

1. *Distintas concepciones del internismo/externismo psicológico*

De acuerdo con el plan establecido, dedicaremos esta sección a distinguir y analizar una serie de opciones teóricas que se han atribuido a sí mismas la denominación de “internismo” o “externismo” psicológico pero que, a pesar de coincidir en su denominación, se ocupan de aspectos considerablemente diferentes de nuestros procesos mentales y pensamientos. Es conveniente realizar esta tarea de análisis por el siguiente motivo. Aun cuando desde el comienzo de nuestro trabajo se ha destacado que la variante de externismo psicológico que presuntamente entraba en conflicto con el acceso privilegiado era únicamente el externismo relativo a la individuación de nuestros contenidos mentales, en algunos puntos de la exposición han intervenido en la polémica autores que defendían (o atacaban) otras concepciones del externismo. Por esta razón, proponemos sistematizar de la manera más exhaustiva posible dichas variantes.

(1) Internismo y externismo con respecto a la individuación (o internismo/externismo constitutivo):

Estos enfoques teóricos se perfilan a partir de las dos respuestas posibles a la siguiente cuestión: la *identidad* de (al menos) algunos de nuestros *procesos cognitivos* y de (al menos) algunos de nuestros *contenidos mentales*, ¿depende de factores externos al hablante? Una respuesta negativa a esta pregunta supone respaldar el internismo constitutivo, mientras que una respuesta afirmativa es propia de las opciones externistas en este sentido.

Además, dentro del internismo/externismo constitutivo, es preciso distinguir dos variantes, dependiendo de si son los procesos cognitivos o los conceptos/contenidos aquello cuya dependencia con respecto al entorno externo se discute (Menary 2010, pp. 1-3):

(1.a) Internismo y externismo con respecto a la individuación de procesos (también conocido como internismo/externismo constitutivo activo):

La pregunta relevante en este caso es la de si nuestros *procesos* mentales están, al menos parcialmente, constituidos por el entorno externo y si, en este sentido, son dependientes del mismo. Al responder de manera negativa estaremos manteniendo una opción internista, de acuerdo con la cual ninguno de estos procesos (memoria, percepción, cálculo lógico-matemático, etc.) viene constituido, ni siquiera en parte, por elementos externos al hablante. En caso de

ofrecer una respuesta afirmativa a dicha pregunta, consideraremos que la constitución ontológica de los procesos mentales involucra elementos del mundo externo (cuadernos, telescopios, calculadoras, ordenadores) como parte integrante de dichos procesos. Los partidarios de esta última opción son conocidos como partidarios de la “teoría de la mente extendida”.

(1.b) Internismo y externismo con respecto a la individuación de contenidos (también conocido como internismo/externismo constitutivo pasivo):

Esta es la clase de externismo cuya compatibilidad con el acceso privilegiado se pone en cuestión. Como ya vimos, la pregunta que los externistas e internistas discuten en este ámbito es la siguiente: la identidad de (al menos) algunos de los contenidos y conceptos de un hablante, ¿depende exclusivamente de las propiedades internas de este sujeto? En caso de que nos decantáramos por una respuesta afirmativa a esta pregunta, estaremos respaldando el internismo constitutivo pasivo. Si sostuviéramos que, al menos en ocasiones, es necesario contar con elementos externos al hablante, estaríamos adhiriéndonos al externismo en este sentido.

Una vez descrita la oposición entre internismo/externismo en relación a la individuación de procesos y de contenidos, conviene explicar en qué sentido el primero de ellos recibe la denominación de “(externismo constitutivo) *activo*” y por qué el segundo es calificado de *pasivo*. El matiz radica en el hecho de que para el internismo/externismo constitutivo pasivo, qué contenidos y conceptos expresa un hablante depende de qué factores (internos o parcialmente externos) *han figurado* de manera predominante en el perfil causal de dicho sujeto. Esto es, la *historia* de interacciones de un hablante (con géneros naturales de su entorno, con las convenciones lingüísticas de su comunidad) es lo que determina para la opción externista qué contenidos piensa este sujeto *en el presente*. En cambio, lo que se defiende desde el externismo constitutivo activo es que hay elementos del entorno externo que intervienen activamente en los procesos cognitivos de los hablantes. De esta manera, *cada vez* que tenga lugar un proceso de esta clase habrá que tomar en consideración qué factores externos son los que, en cada caso, están contribuyendo (activamente) al desempeño del proceso *al mismo tiempo* que éste acontece.

Según se ha indicado, en las siguientes secciones de nuestro trabajo profundizaremos en las distintas concepciones del internismo/externismo constitutivo

pasivo o, como lo hemos denominado, internismo/externismo con respecto a la individuación de nuestros contenidos mentales.

(2) Internismo y externismo adquisitivo:

El hecho de optar por el internismo o el externismo en cuanto a la *adquisición de conceptos y contenidos mentales* depende de la respuesta que se dé a la siguiente cuestión: ¿un hablante puede tener conceptos y contenidos sin establecer contacto con entidades externas? De respaldar la primera opción y afirmar, con ello, que podríamos tener conceptos y contenidos mentales aunque no hubiéramos establecido ningún contacto con objetos externos, optaremos por una respuesta internista en este sentido. De lo contrario respaldaremos una opción externista, fuerte o débil. Distingamos entre las siguientes posibilidades:

(2.a) Internismo adquisitivo:

Es posible que un sujeto posea todos sus conceptos o contenidos mentales sin que haya establecido ningún contacto con entidades del mundo externo. Esto se debe o bien a que esos conceptos/contenidos son innatos o bien a que son conceptos/contenidos construidos a partir de conceptos/contenidos innatos.

(2.b) Externismo adquisitivo fuerte:

Nuestros conceptos/contenidos sólo pueden tener un origen externo, lo cual supone que un sujeto carecerá de conceptos/contenidos a menos que interactúe con objetos del mundo externo.

(2.c) Externismo adquisitivo débil:

Un sujeto podría tener algunos conceptos/contenidos sin haber establecido contacto con los objetos externos.

Como puede verse, cada una de estas opciones es una tesis modal sobre las *condiciones de adquisición* de conceptos y contenidos mentales. Tanto el internismo adquisitivo como el externismo adquisitivo débil niegan que el contacto con el mundo externo constituya una condición necesaria para adquirir al menos algunos de nuestros conceptos/contenidos (para todos ellos, en el caso del internismo adquisitivo). El externismo adquisitivo fuerte, por su parte, defiende el origen externo de todos nuestros conceptos y contenidos, por lo que el contacto con los objetos externos sería una condición necesaria para la adquisición conceptual. Por otro lado, sólo esta última opción, el externismo adquisitivo fuerte, es imposible de compatibilizar con el

innatismo conceptual, es decir, con la teoría filosófica que sostiene que (al menos) algunos de los conceptos/contenidos que poseemos no provienen de la interacción con aspectos externos del mundo. El externismo adquisitivo débil no se compromete con la idea de que algunos de los conceptos y contenidos que *de hecho* poseemos sean innatos, sino que simplemente se limita a conceder que podríamos tener conceptos y contenidos de esa clase. Por último, el internismo adquisitivo conlleva un compromiso claro con el innatismo conceptual, ya que si se defiende que los conceptos que posee un hablante no tienen un origen externo, sólo cabe mantener que este hablante los posee de manera innata o los ha construido a partir de otros conceptos que posee de esa manera.

Según comprobamos en el capítulo anterior, algunos partidarios de la individuación externista de nuestros contenidos, es decir, de la tesis que propiamente entra en conflicto con el acceso privilegiado a los propios contenidos mentales, han confundido el internismo con respecto a la individuación con el internismo de adquisición (o innatismo conceptual).²⁷⁷ No obstante, la cuestión de la adquisición conceptual y la cuestión de la individuación conceptual son distintas y admiten respuestas independientes. De hecho, existen autores como Searle que han mantenido un externismo adquisitivo fuerte, de acuerdo con el cual un sujeto que no hubiera establecido interacciones con objetos externos no poseería concepto alguno y, sin embargo, han defendido al mismo tiempo que las propiedades internas de este sujeto determinan por sí mismas qué conceptos posee en un momento dado.²⁷⁸ Dicho de otra manera, es posible conjugar coherentemente estas dos tesis, a saber, que si un agente jamás hubiera entrado en contacto con objetos externos no poseería ningún contenido mental y, en segundo lugar, que una vez que ha adquirido tales contenidos cualquier variación en el entorno externo del agente que deje inalteradas sus propiedades internas no afectará a la identidad de sus contenidos mentales.

(3) Internismo y externismo descriptivo:

Estas opciones teóricas se establecen a partir de las dos posibles respuestas a la siguiente cuestión: ¿cabe *especificar* los *conceptos y contenidos* mentales de un hablante empleando un vocabulario que no haga referencia a objetos externos a este

²⁷⁷ Un ejemplo de esto lo encontramos en Sawyer (1998), p. 531.

²⁷⁸ Jackson adopta la misma posición en (2003, p. 56) al afirmar que “la determinación del contenido, tanto semántico como intencional, involucra, en algún sentido amplio, relaciones causales entre el entorno y el sujeto. Pero esto es consistente con la afirmación de que el contenido (representacional) es estrecho en tanto que no depende del entorno, en el sentido de que no es una propiedad que podamos cambiar simplemente con un transporte de un entorno a otro”.

sujeto? En caso de responder afirmativamente, se respaldará la opción internista; de negarlo, cabe adherirse a una posición externista fuerte o débil. Las opciones lógicas frente a esta pregunta son las siguientes:

(3.a) Internismo descriptivo:

Todos nuestros contenidos mentales pueden ser especificados mediante un lenguaje cuyos términos no hacen referencia a objetos externos a nosotros mismos. Se defiende, por tanto, que es posible usar un lenguaje que se refiera exclusivamente a datos sensibles u otra clase de entidades puramente internas para caracterizar todos nuestros conceptos y contenidos.

(3.b) Externismo descriptivo fuerte:

Ninguno de nuestros conceptos y contenidos puede ser especificado mediante un lenguaje que no se refiera a objetos y estados de cosas externos a nosotros mismos.

(3.c) Externismo descriptivo débil:

Es posible especificar al menos algunos de nuestros conceptos usando un lenguaje que no remita a objetos y estados de cosas externos a nosotros mismos.

Al igual que las opciones (2.a), (2.b) y (2.c), las tres tesis que acabamos de presentar tienen un carácter modal, pronunciándose, esta vez, sobre la posibilidad de describir los conceptos y contenidos mentales de un hablante sin hacer referencia a objetos externos a éste. Desde luego, existe una cuestión previa al posicionamiento internismo-externismo descriptivo y es la de si es posible un lenguaje cuyos términos no hagan ninguna referencia a objetos externos a los hablantes.

De la misma manera que en el caso de las opciones anteriores, en los capítulos precedentes hemos comprobado que algunos participantes en la polémica sobre la presunta incompatibilidad entre el acceso privilegiado y el externismo con respecto a la individuación confundían esta última tesis con el externismo descriptivo fuerte. En este sentido, en la tercera sección del segundo capítulo tomamos en consideración la posición de Davidson, quien, por un lado, declaraba que no existe conflicto alguno entre el acceso privilegiado y el hecho de atender a factores externos para describir los pensamientos de los hablantes pero, por otro lado, defendía que los estados mentales dotados de contenido de un agente son metafísicamente idénticos a sus estados neuronales y, por tanto, los concebía como propiedades internas. Esto indicaba, como

vimos, que Davidson, en realidad, no respaldaba un genuino externismo con respecto a la individuación (Davidson 1987, p. 67 y ss.; 1988a, pp. 85 y ss.).²⁷⁹

Al hilo de esta distinción es conveniente examinar un argumento confeccionado por Bilgrami en (1992a) con el propósito de demostrar que los partidarios del internismo con respecto a la individuación – especialmente Searle – se encuentran comprometidos con un internismo descriptivo, tesis que, en su opinión, resulta extremadamente implausible. La objeción de Bilgrami queda recogida en la siguiente pregunta retórica: "¿Por qué [la posición internista] debe usar irremediamente un vocabulario conceptual, cuyo punto preciso es referirse al mundo externo en la especificación de contenidos, cuando la clave principal de su posición es, precisamente, que los contenidos no están determinados por nada en el mundo externo?" (Bilgrami 1992a, p. 221). De acuerdo con este autor, un partidario del internismo con respecto a la individuación como Searle no tiene derecho a usar un vocabulario que remita a objetos externos con el fin de especificar los estados mentales de un hablante puesto que, según su propia posición, dichos estados mentales son *independientes* de cualquier objeto externo. Sin embargo, pensamos que la acusación de Bilgrami no es del todo apropiada, especialmente porque su definición de "internismo con respecto a la individuación" resulta muy imprecisa ("los contenidos no están determinados por nada externo"), al no especificar la clase de independencia entre nuestros conceptos/contenidos y los factores externos que defiende un internista de este tipo.

Por otro lado, Bilgrami no sólo exhibe una comprensión excesivamente vaga de la "independencia" que, según el internismo con respecto a la individuación, existe entre nuestros contenidos y el mundo externo, sino que, además, cabe interpretar su argumento como si se sustentara en una teoría de la referencia que difícilmente sería respaldada por un partidario de esa clase de internismo. Para explicar esta idea, distingamos, por un lado, la cuestión de si en los informes psicológicos usamos un vocabulario que *remite* a entidades externas y, por otro, la pregunta de si las propiedades semánticas de sus términos son *dependientes* de tales entidades (externismo semántico). La objeción de Bilgrami tendría sentido si estas dos propiedades de un vocabulario fueran equivalentes, pero existen buenas razones para pensar que un partidario del internismo psicológico no tiene por qué considerar legítima esta equivalencia. Es más, desde el internismo hay recursos para defender que nuestros

²⁷⁹ Cf. Owens (2003), pp. 204-205.

términos hacen referencia a entidades externas al hablante sin comprometerse con el externismo semántico. Bastaría con que la propia ubicación interna de las propiedades semánticas de nuestro vocabulario y de nuestros contenidos mentales no sólo no impidiera, sino que constituyera la condición necesaria para llevar a cabo esa remisión o referencia a entidades externas.

Vamos a explicar con cierto detenimiento nuestra objeción a Bilgrami. De acuerdo con este autor, no es posible emplear un lenguaje que remita a entidades externas para describir estados mentales constituidos independientemente de tales entidades. Sin embargo, mantenemos que nada impediría lograr tal cosa siempre y cuando las propiedades semánticas de ese lenguaje también quedaran internamente constituidas. Con otras palabras, el problema sólo surge cuando pretendemos describir estados mentales individuados de manera puramente interna usando un lenguaje que funciona según las tesis del externismo semántico (un lenguaje cuyas propiedades semánticas están individuadas, en parte, por factores del entorno externo a los hablantes). En efecto, quienquiera que mantenga esta última posición tendrá que hacer frente al argumento desarrollado por Sawyer en (2007, p. 27-28) cuya conclusión es que nuestros estados mentales quedarían, en tal caso, relegados a la infabilidad.²⁸⁰

Para comprobarlo, enunciemos dicho argumento como un trilema que intenta demostrar la incompatibilidad de estas tres tesis:

- (a) Las propiedades semánticas de la mayor parte de nuestros términos lingüísticos se determinan de manera externista.
- (b) Nuestros pensamientos tienen un contenido individuado de manera exclusivamente interna.
- (c) Es posible ofrecer una especificación lingüística del contenido de nuestros pensamientos.

Puesto que la tesis (c) se encuentra firmemente respaldada por nuestras prácticas cotidianas de atribución psicológica, los miembros del trilema bajo sospecha son (a) y (b).

Con el fin de apreciar en qué punto reside la tensión entre estas tres tesis, retomemos el experimento mental de Putnam sobre los términos “aluminio” y “molibdeno” y supongamos, como pide (a), que la mayor parte de los términos de nuestro lenguaje, incluidos los términos de género natural como “aluminio” y

²⁸⁰ Un argumento similar puede encontrarse en Rowlands (2003), pp. 114 y ss.

“molibdeno”, tienen un comportamiento semántico externista. Esto supone que dos hablantes pueden asociar con un mismo término un conjunto de propiedades cualitativamente indistinguibles, sin perjuicio de que dicho término posea un significado distinto, según lo use uno u otro, en función de factores del entorno externo a dichos hablantes (propiedades subyacentes de los géneros naturales y/o convenciones sociales que rigen el uso correcto del término). La consecuencia que se desprende de mantener (a) es, en términos de Sawyer, que nuestro vocabulario no es *semánticamente neutro*, es decir, que las propiedades semánticas de nuestro lenguaje dependen, en parte, de nuestro entorno externo. Por otro lado, de acuerdo con (c), parece intuitivo mantener que a menudo usamos nuestro lenguaje para elaborar descripciones de los contenidos mentales propios y ajenos. Ahora bien, el problema aparece cuando nos adherimos al internismo con respecto a la individuación, recogido en el miembro (b) del trilema. De este modo, si uno mantiene que nuestros contenidos mentales están individuados de manera puramente interna, entonces los hablantes que protagonizan el experimento mental del aluminio y el molibdeno se encontrarán, en este caso, en un mismo estado psicológico. A la hora de elaborar una especificación lingüística de ese estado, podría proponerse una descripción del tipo “ambos hablantes piensan que los enseres de cocina están fabricados con *el material brillante usado para fabricar las bicicletas*”. En este caso, la descripción destacada en cursiva recogería aquella propiedad cualitativamente idéntica que ambos agentes asocian con la extensión del término “aluminio”.²⁸¹ Pero, de acuerdo con el externismo semántico enunciado en la premisa (a) del trilema, esta especificación lingüística del pensamiento de ambos hablantes no es semánticamente neutra o, lo que es igual, cabe pensar que el significado de los términos involucrados en dicha especificación recibe, a su vez, una individuación externista. Para comprobar esto, podemos concebir una prolongación del experimento mental del aluminio y el molibdeno introduciendo un segundo escenario contrafáctico. En este caso, un tercer hablante, internamente idéntico a los hablantes iniciales, profiere la oración “los enseres de cocina están fabricados con aluminio”. Supongamos que en el entorno social en el que está inserto este tercer agente el término “bicicleta” tiene asociados unos usos lingüísticos diferentes de los que están asociados a ese mismo término en los entornos de los dos hablantes iniciales. Imaginemos, por ejemplo, que además de designar los vehículos de dos ruedas que nosotros conocemos con ese nombre, el término “bicicleta”

²⁸¹ Para simplificar supongamos que ambos asocian una única propiedad de esta clase. Esto no modificará en lo esencial el argumento, como puede comprobarse fácilmente.

también se aplica a los balones hinchables de color azul. Sin embargo, concedamos que este tercer sujeto desconoce que los balones de estas características también se denominan en su entorno “bicicletas” y que, además, asocia con el término “aluminio” las mismas propiedades cualitativas que los dos hablantes iniciales. Lo interesante de esta situación es que la diferencia que existe en los usos lingüísticos asociados al término relevante (“bicicleta”) nos impide utilizar la descripción lingüística “estos hablantes piensan que los enseres de cocina están fabricados con *el material brillante usado para fabricar las bicicletas*” para caracterizar el estado mental relevante (el mismo, de acuerdo con el internismo con respecto a la individuación). La razón es que, según las pautas del externismo (social) con respecto a la individuación, el concepto que expresa nuestro tercer hablante mediante el término “bicicleta” es distinto del concepto que expresan los dos hablantes iniciales haciendo uso del mismo término.

A partir de esto, Sawyer sostiene que de mantener un externismo semántico robusto como el recogido en la premisa (a) del trilema, cabe repetir indefinidamente el procedimiento ejemplificado y estipular, así, una variación contrafáctica en el significado de cada uno de los términos que figuran en la presunta especificación lingüística del contenido mental individuado de manera puramente interna. De aceptar que nuestro lenguaje no es semánticamente neutro (i.e. que la semántica de sus términos recibe una individuación externista) entonces, en efecto, existen serias dificultades para ofrecer especificaciones lingüísticas de contenidos mentales individuados de una manera puramente interna.

Ahora bien, cabe alegar que la combinación de las tesis (a) y (b) del trilema de Sawyer, es decir, un externismo semántico fuerte, por un lado, y un internismo con respecto a la individuación, por otro, no es una posición teórica que cuente con demasiado respaldo. De hecho, el propio Searle, a quien Bilgrami cita como interlocutor polémico, no es un partidario de la primera de estas tesis. Como se ha dicho, un internista *à la* Searle consideraría que la determinación internista de las propiedades semánticas de nuestro lenguaje no conlleva despojar a nuestro vocabulario de la capacidad de remitir a entidades externas. Lo que sostendría es, por una parte, que el hablante que usa un término general para referirse con éxito a una clase de entidades ha de tener determinadas propiedades internas y, por otra parte, que dos sujetos cuyas propiedades internas son exactamente idénticas estarán refiriéndose a las mismas entidades externas. Con otras palabras, un defensor del internismo en semántica *no rechaza que nuestro lenguaje remita a entidades externas*, sino que establece como

requisito para dicha remisión que los hablantes posean determinadas propiedades internas. Las declaraciones de Bilgrami, en suma, resultan bastante cuestionables, ya que no parece existir nada de paradójico en aceptar una semántica internista y usar, por ello, un vocabulario que remite (de forma mediata) a entidades externas, en conjunción con una teoría de la individuación de nuestros contenidos que hace de la remisión a entidades externas un proceso internamente mediado.

2. *Problemas de la clasificación usual de las variedades del externismo con respecto a la individuación*

Una vez distinguidas las distintas tesis que se agrupan bajo el rótulo de “internismo” o “externismo” psicológico, en la presente sección abordaremos tres cuestiones. En primer lugar, expondremos la clasificación más usual de las diferentes variedades de externismo sobre la individuación de nuestros contenidos mentales. En una segunda subsección examinaremos una serie de fenómenos psico-lingüísticos, distintos de los experimentos mentales mencionados, en los que encuentran apoyo las distintas tesis externistas. En tercer lugar, se argumentará que la tipología usual expuesta al comienzo de esta sección no es todo lo rigurosa que sería deseable. Según creemos, en su elaboración se ha pasado por alto un hecho de fundamental importancia, a saber, que en ciertos casos la individuación en función de factores naturales determina una identidad para nuestros conceptos *distinta* de la identidad que determinan al respecto los factores externos de carácter social. Por este motivo, cualquier externista que pretenda defender que *tanto* los elementos sociales *como* los elementos naturales desempeñan un papel en la constitución de nuestros conceptos, habrá de realizar un trabajo adicional en la elaboración de su tesis, con el objetivo de evitar los posibles conflictos en la influencia de ambos tipos de factores.

2.1 *Tipología usual de las variedades de externismo con respecto a la individuación de contenidos mentales*

Siguiendo con lo indicado, dedicamos esta subsección a presentar la clasificación al uso de las distintas concepciones del externismo sobre la individuación de nuestros estados mentales. De esta manera, si la primera sección se reservó para el examen de las distintas tesis que se presentan a sí mismas bajo el nombre de “externismo”, a pesar de

que tratan cuestiones muy distintas en relación a nuestros estados mentales, el externismo acerca de la individuación es un enfoque que, a su vez, admite distintas variantes. Al comienzo de este trabajo ofrecimos una definición de mínimos de esta tesis, en tanto que se trata de una concepción que nos invita a reconocer la influencia que los factores externos de tipo “natural y/o social” ejercen en la individuación de nuestros contenidos mentales. En esta subsección presentaremos un cuadro en el que se despliegan las diferentes instancias de esta descripción general, atendiendo especialmente a la dependencia que se presume entre los contenidos de nuestros pensamientos y (cada una de) estas dos clases de factores.

No obstante, antes de desarrollar la tipología usual de las variedades de externismo conviene recordar qué concepción de los géneros naturales y de la referencia de los términos de género natural subyace a las concepciones del externismo que reconocen la influencia del entorno natural. Aunque este asunto ha sido tratado en varias secciones de nuestro trabajo, debemos introducir algunas precisiones que serán de utilidad en lo siguiente. La teoría de la referencia de los términos de género natural que respaldan aquellos externistas que reconocen la influencia del entorno externo natural en la individuación de los contenidos mentales es la teoría expuesta por Putnam principalmente en (1975c),²⁸² teoría que, según comprobamos en el primer capítulo de nuestro trabajo, guarda un claro parentesco con la defendida por Kripke al respecto en (1980).²⁸³ Ambos autores consideran que los géneros naturales incluyen un subconjunto

²⁸² También en Putnam (1970) y (1973b). No obstante, debe señalarse que la mayor parte de este último artículo, incluyendo los experimentos mentales de la Tierra Gemela, forma parte de (1975c).

²⁸³ Conviene señalar que aunque existen muchos paralelismos entre las teorías de la referencia de los términos de género natural mantenidas por ambos autores, ello no significa que sus opiniones sean perfectamente coincidentes. De este modo, existen dos aspectos importantes en los que difieren sus teorías, a saber, (i) el papel de los expertos en la determinación de la referencia de los términos de género natural y (ii) la necesidad (o no) de que un hablante tenga al menos ciertas creencias verdaderas acerca de los ejemplares que pertenecen a un género natural para poder referirse con éxito a tales ejemplares mediante el uso del término relevante. En relación a (i), Putnam considera que para que un hablante no-experto pueda usar con éxito un término de género natural es más importante el vínculo existente entre este individuo y los hablantes expertos de su comunidad lingüística que la inserción del primero en una cadena histórica de usos del término que le relacione causalmente con la introducción inicial de la expresión en su lenguaje. Para Kripke, sin embargo, la prioridad entre ambos factores es justamente la inversa. El motivo de discrepancia recogido en (ii) se pone de manifiesto si tomamos en consideración las observaciones de Putnam (1973a, p. 203) sobre la referencia de los nombres propios y las extendemos al ámbito de los términos de género natural. En este sentido Putnam afirma que él no es tan “liberal” como Kripke con respecto al conocimiento que cabe exigir a un hablante para que pueda referirse con éxito al portador de un nombre propio haciendo uso de dicho nombre. Acudiendo al ejemplo que él mismo nos propone, un hablante que únicamente asocie con el nombre “Quine” la creencia de que fue un emperador romano, no puede referirse a Quine cuando haga uso de ese nombre. Sin embargo, para Kripke este hablante puede referirse a Quine siempre y cuando se encuentre inserto en la cadena de transmisión histórico-causal de dicho nombre. A este respecto sostenemos que cabe extender de manera plausible

de entidades denominadas *ejemplares paradigmáticos*. El estatus de estos ejemplares – en tanto que paradigmáticos – depende del hecho de que tales entidades poseen ciertas propiedades que consideramos típicas o características de los miembros del género, propiedades a las que Putnam se referirá como *propiedades distintivas normales* (1970, p. 139) y que conciernen a la apariencia externa de dichas entidades. No obstante, tanto Kripke como Putnam coinciden en señalar que, además de esta clase de propiedades, los miembros de un determinado género natural poseen, en cuanto tales, una misma naturaleza, determinada por el hecho de que todos ellos comparten idéntica *estructura* (Kripke 1980, pp. 120 y ss.) o una(s) misma(s) *propiedad(es) subyacente(s)* (Putnam 1970, p. 139). La tesis fundamental de estos autores acerca de la referencia de los términos de género natural consiste en que la extensión de esta clase de términos viene determinada por la posesión de esa estructura o propiedad(es) subyacente(s) por parte de los ejemplares del género.²⁸⁴ De este modo, esta clase de propiedades constituye, primeramente, el criterio de identidad de un género natural, de acuerdo con lo cual en relación al género natural K, sólo las entidades que poseen las propiedades subyacentes en cuestión son del género K.²⁸⁵ A este respecto cabe señalar, en primer lugar, que tanto Kripke como Putnam se han referido a las propiedades subyacentes como propiedades *esenciales*, es decir, como propiedades que constituyen lo que es ser (un) K y no cualquier otro tipo de cosa.²⁸⁶ En segundo lugar, en contraste con las propiedades distintivas normales que caracterizan a los miembros paradigmáticos de un género natural y que conciernen a la apariencia externa de los ejemplares del género, las propiedades subyacentes son *internas* y *causalmente responsables* de las propiedades distintivas normales y, además, *objeto de descubrimiento científico*.

Tras esta breve recapitulación de los aspectos más relevantes de esta teoría de la referencia de los términos de género natural, conviene realizar una observación de cierta importancia. En lo sucesivo tomaremos estos puntos de coincidencia entre las teorías de Kripke y de Putnam como *la teoría de la referencia de los términos de género natural*

estas observaciones de Putnam acerca de la referencia de los nombres propios a la referencia de los términos de género natural. Cf. Hacking (2007).

²⁸⁴ En lo que sigue emplearemos de manera equivalente las expresiones “propiedad(es) subyacente(s)” y “*estructura subyacente*”.

²⁸⁵ Las propiedades subyacentes que poseen los miembros de un género natural son, asimismo, el criterio de identidad inter-mundano del género. De este modo, en un mundo posible habrá (ejemplares del género natural) K si y sólo si en ese mundo hay individuos que posean la propiedad subyacente relevante.

²⁸⁶ Así, tanto Kripke en (1980), pp. 126, 133, 135 y 138 como Putnam en (1970), p. 140, hablan de una “naturaleza esencial” compartida por todos los ejemplares de un género natural.

asumida por los partidarios del externismo natural.²⁸⁷ Sin embargo, es preciso reconocer dos hechos. En primer lugar, el apoyo que Putnam prestó a las tesis esencialistas en sus escritos posteriores a (1975c) se ha ido reduciendo considerablemente, hasta el punto de que ha llegado a mantener que la “esencia” (término que el propio Putnam pone entre comillas) no es tanto una propiedad que constituye la condición necesaria y suficiente de pertenencia a un género natural, como *un tipo de paradigma* al que deben asemejarse otras aplicaciones del concepto correspondiente (Putnam 1983a, p. 64). En cambio, los partidarios del externismo que reconocen la influencia del entorno natural en la individuación de nuestros contenidos no han atendido a estas consideraciones posteriores de Putnam. En segundo lugar, existen evidencias en el propio texto de (1975c) que reflejan que la teoría de la referencia de los términos de género natural que Putnam propone allí, así como la concepción de los géneros naturales que subyace a dicha teoría, son, en realidad, concepciones mucho menos monolíticas de lo que usualmente han pretendido los partidarios esta clase de externismo. Tal y como ha destacado Fernández Moreno en (2000), la sección de Putnam (1975c) que lleva por título “Otros sentidos” contiene observaciones que revelan cierta flexibilidad en sus propuestas, hasta el punto de que parecen obligarnos a revisar si lo que suele presentarse como *la* teoría de la referencia de los géneros naturales de Putnam es, en verdad, la teoría que este autor propuso. Con el fin de justificar esta afirmación, comprobemos que Putnam sostiene que para que algo sea agua ha de compartir con los ejemplares paradigmáticos de agua ciertas propiedades físicas importantes (algo a lo que ya hemos aludido en las dos páginas anteriores). Lo llamativo es la explicación que nos ofrece a continuación sobre la importancia de dichas propiedades:

“La *importancia* es una noción relativa a los *intereses*. Normalmente, las propiedades ‘importantes’ de un líquido, un sólido, etc., son aquellas que resultan estructuralmente importantes: aquellas que especifican de qué está hecho el líquido, sólido, etc. – partículas elementales, hidrógeno y oxígeno, de tierra, aire, fuego, agua, o lo que sea – y cómo están unidas o combinadas para producir las características superficiales [...] [Pero] en un contexto ‘agua’ puede significar agua químicamente pura, mientras que en otro contexto puede significar la sustancia del lago Michigan. Y *un hablante puede en ocasiones*

²⁸⁷ Dejando a un lado, como es natural, aquellos puntos de discrepancia entre ambos autores que señalamos en la nota 283.

referirse al XYZ como agua si uno está usándolo como agua” (Putnam 1975c, p. 239; la cursiva es nuestra).

Este fragmento resulta extremadamente revelador, puesto que en él Putnam admite que, por lo general (no siempre), las propiedades subyacentes determinan la pertenencia a un género natural, así como la extensión del término de género natural correspondiente, *en la medida en que dichas propiedades son normalmente las importantes desde el punto de vista de los intereses de los hablantes*. De esto se sigue, en última instancia, que son nuestros intereses los que señalan qué tipo de propiedad es aquella cuya posesión por parte de las entidades es la determinante para su pertenencia a un género natural. Desde luego, resulta sorprendente que Putnam admita, además, que un hablante puede referirse al XYZ como agua, cuando el fulcro del experimento mental presentado en ese trabajo consiste, precisamente, en negar de manera taxativa que dicho líquido fuera agua. Estas tesis, por tanto, parecen reflejar una cierta tensión entre las ideas de Putnam. Conjeturamos que el hecho de reservar esta posición tan destacada a los intereses de los hablantes constituye un indicio de la evolución posterior de su pensamiento que, como señalamos, se aleja del esencialismo. En cualquier caso, el Putnam más flexible de la sección titulada “Otros sentidos” ha tenido menos influencia entre los partidarios de las tesis del externismo con respecto a la individuación que el Putnam esencialista y científicista.²⁸⁸ Debido a este hecho, vamos a tomar en consideración en nuestro trabajo únicamente la concepción de la referencia de los términos de género natural de Putnam que ha tenido verdadera influencia, que es la concepción desarrollada en Putnam (1970), (1973b) y, sobre todo, (1975c) y que, como señalamos, guarda importantes similitudes con la concepción kripkeana.

Una vez realizadas estas aclaraciones sobre la teoría de la referencia de los términos de género natural que subyace a algunos enfoques externistas, ofrecemos a continuación la tipología más usual de las distintas concepciones del externismo, en función de cuál sea la influencia que se reserve a los factores de carácter natural y social en la individuación de nuestros contenidos mentales (Brown 2004, pp. 9-20):

²⁸⁸ En este sentido, Putnam desarrolla su teoría de la referencia de los términos de género natural, además de en (1975c), en el capítulo 4 de (1983b), en algunas de las secciones del capítulo 2 de (1988), en el capítulo 4 de (1990) y en el capítulo 3 de (1994). Sin embargo, aunque la mayoría de los artículos relevantes para el debate sobre la compatibilidad del externismo y el acceso privilegiado se escribieron con posterioridad a la publicación de estos artículos y libros de Putnam, el único trabajo de este autor que la mayoría de los partidarios y detractores del externismo psicológico toman en consideración es, a lo sumo, (1975c).

Externismo natural: La identidad de los contenidos mentales que un hablante expresa usando *términos de género natural* depende, en parte pero fundamentalmente,²⁸⁹ de las propiedades subyacentes de los ejemplares que pertenecen a la extensión de tales términos, concebidas según la teoría de los géneros naturales de Putnam. Esta tesis se concreta de la siguiente manera siguiente: dos hablantes exactamente idénticos en todos sus aspectos internos pero pertenecientes a entornos distintos expresarán conceptos diferentes al usar el término de género natural “T” si los ejemplares que pertenecen a la extensión de “T” poseen, en cada uno de sus entornos, unas propiedades subyacentes distintas. Esta diferencia existe incluso aunque los ejemplares que reciben el nombre de “T” en cada entorno sean indistinguibles para cualquier hablante de cualquiera de estos entornos (i.e. aunque nadie conozca esa diferencia en las propiedades subyacentes de los miembros de los géneros).

Externismo social: La identidad de los contenidos mentales que un hablante expresa usando un *término general (de género natural o no)* depende, en parte pero fundamentalmente, del uso que la comunidad lingüística de dicho hablante ha establecido para dicho término. Esta tesis se concreta de la siguiente manera: dos sujetos exactamente idénticos en todos sus aspectos internos que están insertos, cada uno de ellos, en comunidades lingüísticas que establecen convenciones diferentes para el uso del término “T” expresarán conceptos distintos mediante el uso de dicho término. Esta diferencia existe incluso aunque tales sujetos (en particular) desconozcan la diferencia en las convenciones lingüísticas comunitarias que rigen el uso del término en un entorno y en el otro.

Externismo amplio o mixto: La identidad de los contenidos mentales que un hablante expresa usando un *término general (de género natural o no)* depende, en parte pero fundamentalmente, del uso que la comunidad lingüística de dicho hablante ha establecido para dicho término y de las propiedades subyacentes de

²⁸⁹ Con la expresión “en parte pero fundamentalmente” queremos dar a entender que los factores externos que, según el externista, determinan la identidad de nuestros contenidos mentales no son los únicos factores que han de tomarse en consideración para la individuación de dichos contenidos. Sin embargo, una diferencia en los factores externos relevantes será condición suficiente para establecer una diferencia en los contenidos mentales correspondientes.

los ejemplares que pertenecen a los géneros naturales (concebidas según la teoría de los géneros naturales de Putnam) si se trata de un término de género natural. Esta tesis se concreta de la siguiente manera: dos sujetos exactamente idénticos en todos sus aspectos internos que están insertos, cada uno de ellos, o bien en comunidades lingüísticas que establecen convenciones diferentes para el uso del término “T” o bien en entornos distintos donde los ejemplares de T poseen, respectivamente, diferentes propiedades subyacentes, expresarán conceptos distintos mediante el uso de dicho término. Esta diferencia existe incluso aunque tales sujetos desconozcan la diferencia en las convenciones lingüísticas comunitarias que rigen el uso del término o la diferencia en las propiedades subyacentes de los ejemplares que en cada entorno se denominan “T”.

Hechas estas distinciones, vamos a recordar, primeramente, qué clase de experimentos mentales han respaldado cada una de estas variantes de externismo. En primer lugar, el experimento mental de la Tierra Gemela de Putnam en torno al uso del término “agua” constituye el sustento teórico del externismo natural. En efecto, en el escenario que Putnam nos propone, la existencia de diferentes propiedades subyacentes de las porciones de líquido que se denominan “agua” en dos entornos distintos propicia, según los partidarios del externismo psicológico que han hecho uso de dicho experimento, que los hablantes de cada entorno expresen conceptos distintos cuando usan el término. El hecho que determina que el tipo de externismo respaldado por este experimento sea el externismo natural (y no el externismo social) es, justamente, que ningún hablante de esos entornos conoce esa diferencia en la estructura subyacente de los líquidos respectivos. O, dicho de otra manera, en el escenario que describe este experimento mental no existe la más mínima diferencia en los criterios públicos de uso del término “agua”, supuesto el carácter indistinguible de ambos líquidos por lo que respecta a todos los habitantes de los entornos en cuestión. Al poder que se atribuye a las propiedades subyacentes de los géneros en relación a la individuación de los conceptos de género natural cuando dichas propiedades son desconocidas para todos los hablantes de la comunidad lo denominaremos *influencia natural directa*.

En segundo lugar, los otros dos experimentos mentales que Putnam propone en (1975c) sirven de apoyo al externismo social y al externismo mixto o amplio. Tanto en el caso del par “aluminio”-“molibdeno”, como en el caso del par “olmo”-“haya”, existe, además de una diferencia en las propiedades subyacentes de los géneros naturales

relevantes, una diferencia en las condiciones de aplicación (socialmente determinadas) de estos términos de género natural. Aunque tal diferencia puede resultar desconocida para un hablante particular que usa cualquiera de estos términos, este hecho no impide, según la lectura externista de estos experimentos mentales, que exista una diferencia entre, por un lado, los conceptos de aluminio y de molibdeno y, por otro lado, los conceptos de olmo y de haya que expresa este hablante haciendo uso de los términos correspondientes. En tercer lugar, el experimento mental de Burge en torno al concepto de artritis respalda el externismo social, ya que, según se desprende de sus conclusiones, si un sujeto viviera en un entorno donde las condiciones sociales de aplicación del término “artritis” fueran distintas de las condiciones actuales de aplicación de ese término, entonces los contenidos mentales que expresaría mediante la preferencia de oraciones que involucran ese término también lo serían. De nuevo, el hecho de que ese sujeto en cuestión desconozca la diferencia en las condiciones de aplicación no impide que exista una diferencia relevante entre los contenidos y conceptos expresados.

2.2 Fenómenos psico-lingüísticos que respaldan el externismo con respecto a la individuación

De acuerdo con lo indicado al comienzo de la presente sección, aunque los experimentos mentales mencionados han desempeñado un papel fundamental a la hora de aportar plausibilidad a las tesis externistas, no constituyen el único sustento teórico con el que cuentan dichas tesis. En esta segunda subsección abordaremos de qué modo la perspectiva externista sobre la individuación de los contenidos mentales permite explicar una serie de cuestiones relativas al funcionamiento del lenguaje y a nuestros hábitos de atribución de contenidos.

(1) Fenómeno de la deferencia lingüística

El lenguaje es, en gran medida, una empresa cooperativa. Siguiendo las observaciones de Putnam en (1975c), muchos términos de nuestro lenguaje, es decir, del lenguaje usado por los hablantes de nuestra comunidad lingüística, tienen unas condiciones de aplicación que sólo domina rigurosamente un subconjunto de estos hablantes. Sin embargo, este hecho no impide que el resto de los hablantes podamos usar de manera aceptable esa clase de términos. Así, en la medida en que deleguemos la tarea de

conocer en detalle las condiciones de aplicación de los mismos al subconjunto de expertos relevante, nuestros intercambios lingüísticos pueden transcurrir sin trabas excesivas.²⁹⁰

En un entorno como el nuestro, que presenta un grado tan alto de división del trabajo, con grupos de especialistas dedicados tanto al estudio de diversas áreas de la realidad como a la realización de distintas actividades técnicas, es normal contar con una cantidad ingente de términos cuyas condiciones de aplicación resultan desconocidas por la mayor parte de los hablantes. Existen ejemplos claros como “neutrino”, “lupus” o “conmutador”, pero el mismo fenómeno se presenta en relación a otros términos menos técnicos. Por este motivo, cualquier teoría que otorgue a un sujeto competencia en el uso de un término (y la posesión del concepto relevante) sólo si dicho sujeto conoce las condiciones necesarias y suficientes de la aplicación de ese término será una teoría que calificará de incompetentes a la mayor parte de los hablantes. Ahora bien, aunque el desconocimiento parcial de las reglas de uso de nuestros términos sea de lo más generalizado, resulta evidente tanto el éxito de la mayoría de nuestros intercambios lingüísticos como el hecho de que las incomprensiones están relativamente localizadas y que su resolución suele encontrarse a nuestro alcance. En relación a esto, es digno de destacarse el rol que desempeña la consulta a hablantes expertos en diversas actividades y/o áreas de conocimiento, a diccionarios, manuales u otro tipo de obras escritas por expertos. El externismo social y el externismo mixto constituyen posibles explicaciones conceptuales para este fenómeno y, en esta medida, reciben apoyo del mismo, tal y como puede apreciarse en el propio experimento mental de Burge en torno al uso del término “artritis”. El protagonista del ejemplo desconoce parcialmente el uso del término pero, en la medida en que admite las correcciones de los expertos relevantes de su comunidad (los médicos), el externista interpreta que el concepto que expresa Bert mediante la preferencia de oraciones que involucran el término “artritis” será el mismo concepto que expresan dichos expertos.

Por último, conviene realizar dos aclaraciones. En primer lugar, cuando se ofrece una caracterización de la división social del trabajo lingüístico y del fenómeno de la deferencia sustentado en dicha división, puede parecer que hay, con respecto a cada

²⁹⁰ En este sentido, conviene recordar que para Putnam un hablante posee competencia en el dominio de un término de género natural “T” si asocia con dicho término los marcadores sintácticos, semánticos y los estereotipos apropiados. No es necesario, por tanto, que un individuo conozca la descripción de la extensión que especifica cuál es la propiedad que determina la pertenencia de los individuos a dicha extensión.

término, dos grupos bien diferenciados de hablantes, a saber, el grupo de los expertos, y el resto de los hablantes (o profanos). Sin embargo, esta imagen es un tanto simplista, puesto que la competencia con respecto a un concepto determinado se presenta en grados y, en consecuencia, establecer una división tajante entre expertos y profanos en el seno de una comunidad lingüística no siempre resulta adecuado (Burge 1986c, p. 258). En segundo lugar, en relación a un mismo concepto pueden existir distintos grupos de expertos relevantes en función del trato que uno pretenda establecer con aquello que es designado por el término que expresa el concepto en cuestión. Esto ocurre con respecto a los especialistas en botánica, los agricultores y los cocineros, los cuales constituyen grupos de expertos en relación a determinados términos de género natural (“cebolla”, “tomate”, “coles”, etc.).²⁹¹ Esta idea tiene como consecuencia que cuando hablamos de “condiciones socialmente determinadas para la aplicación de un concepto” estamos refiriéndonos a reglas de uso público para un término que pueden admitir cierta variabilidad en función de cuál sea el subgrupo de expertos que las determine y del momento en el que sean consultados.²⁹² En el caso de términos como “amperímetro” probablemente sólo hay un único grupo de expertos – o uno especialmente destacado –. En cambio, para términos como “oro” puede existir cierta diferencia en las condiciones determinadas por distintos grupos de expertos (químicos, joyeros) o en función del momento histórico en que se lleve a cabo la consulta a los expertos.²⁹³ Lo que debe tenerse en cuenta es que cuando hablamos de las convenciones

²⁹¹ En (1979, p. 388) Dowty emplea uno de estos términos (“tomate”) a modo de ejemplo para defender exactamente lo mismo que nosotros, a saber, que, en ocasiones, un mismo término es usado según distintos estándares, siendo ambos perfectamente legítimos en el seno de un cuerpo social. Merece la pena citar sus palabras al respecto: “Un caso contrario viene proporcionado por [la palabra] “fruta” (en tanto que opuesta a “hortaliza”). Para el botánico, la fruta es un género natural, una parte de una planta que tiene características botánicas comunes en cada especie que pueden ser descubiertas empíricamente. Pero en el uso diario, esta palabra tiene un criterio funcional, i.e. una parte de una planta que es apta para ser comida como postre o tentempié. Me divirtió aprender hace poco (*Scientific American*, Agosto, 1978, p. 78) que la cuestión de si el tomate es una fruta o una hortaliza ha sido dirimida por un cuerpo no menos austero que la Corte Suprema de los Estados Unidos (en 1893). Es una lástima para Kripke y Putnam, pero la Corte Suprema falló contra ellos en este caso. El juez asociado Horace Grey opinó que: *Botánicamente hablando, los tomates son la fruta de una vid, igual que los pepinos, calabacines, judías y guisantes. Pero en el lenguaje común de la gente, tanto de los vendedores como de los consumidores de provisiones, todos ellos son hortalizas que crecen en los jardines domésticos, y, tanto si son comidas crudas o cocidas, son, al igual que las patatas, chirivías, nabos, remolachas, coliflores, coles, apios, y lechugas, servidas usualmente en el almuerzo en, con o después de la sopa, pescado o carne que constituyan la parte principal del menú y no, según ocurre generalmente con las frutas, como postre.*”

²⁹² Aunque, en rigor, las condiciones de aplicación conciernen a los términos que expresan los conceptos relevantes, para abreviar en lo siguiente hablaremos de condiciones de aplicación de dichos conceptos.

²⁹³ “[La división social del trabajo lingüístico], para ser sinceros, es más complicada que una división entre expertos y profanos, puesto que el significado de ‘experto’ depende de las circunstancias. En un caso [con respecto al término ‘oro’] puede ser un joyero y en otro caso, un científico; y sus veredictos no tienen por qué coincidir necesariamente. El científico puede conocer, *qua* científico, que el anillo dorado

sociales asociadas a un término éstas no deben entenderse como un conjunto cerrado de reglas con respecto al cual existe un consenso sin fisuras entre los miembros de un único grupo de expertos relevantes. Aunque en el caso de algunos términos científicos esta imagen puede ser más cercana a la realidad, en otros casos las convenciones distan de ser exhaustivas o no todas ellas son compartidas por todos los grupos de expertos, ya que a menudo tendrán lugar disputas sobre (algunos aspectos de) el significado de los términos y sobre la naturaleza de la clase de cosas designadas mediante la expresión relevante (Burge 1986c, pp. 260).

(2) Parsimonia en el ámbito conceptual

Algunos externistas han señalado que la aplicación de sus tesis a la situación de aquellos hablantes que, al igual que los protagonistas de los experimentos mentales descritos, tienen un conocimiento incompleto de las condiciones de aplicación de sus términos evita la necesidad de postular una infinidad de nociones idiosincrásicas (Woodfield 1998). Esta tesis es una consecuencia de la interpretación externista del fenómeno de la deferencia lingüística ofrecida en (1).

Para comprender esta idea, admitamos provisionalmente el diagnóstico internista acerca de la situación del protagonista del experimento mental en torno al uso del término “artritis”. En este caso, aunque el internista puede admitir que Bert presenta una actitud de deferencia hacia el médico, su tesis conlleva que se ha producido un cambio conceptual en dicho hablante (desde un concepto idiosincrásico, *artritis**, hacia el concepto de artritis) tras su consulta a este experto. Así, el partidario del internismo sostiene que el concepto que dicho sujeto expresa antes y después de su consulta mediante el uso del término “artritis” viene determinado por la información, las creencias y demás actitudes (internas) que él asocia en cada momento con el referente del término “artritis”. Puesto que, por hipótesis, ese cuerpo de propiedades internas es, antes de su conversación con el médico, distinto del cuerpo de propiedades internas que determinan el concepto que expresa el médico usando el mismo término, entonces el protagonista del experimento y el médico expresarán conceptos diferentes, lo que conlleva la necesidad de postular un concepto idiosincrásico de artritis para Bert, distinto del concepto que adquiere tras aceptar la corrección del médico.

que él compró en la joyería no es ‘realmente’ de oro, sin negar que sea realmente de oro en el sentido del joyero” (Hanfling 2000, p. 229).

Ahora bien, los externistas mantienen que el fenómeno de incompreensión parcial de las condiciones de aplicación de un término resulta extremadamente frecuente y, por tanto, si quisiéramos adherirnos a las pautas internistas de individuación tendríamos que repetir este procedimiento para la práctica totalidad de nuestros conceptos. Como hemos dicho, estas pautas, llevadas al límite, conllevarían efectuar una división en el seno de los hablantes de una comunidad lingüística, en relación a aquellos términos cuyas condiciones de aplicación estén sujetas a la división del trabajo lingüístico. Por un lado, encontraríamos a los expertos en el dominio de un determinado término, los cuales expresarían uno y el mismo concepto mediante el uso del término relevante y, por otro lado, el resto de los hablantes o profanos que, al no dominar completamente esas condiciones de aplicación, expresarían una multitud de conceptos idiosincrásicos. Para el externista, este diagnóstico supone atentar contra el principio de parsimonia en el ámbito conceptual y, además, hace de las atribuciones psicológicas una tarea de todo punto inmanejable. En este sentido, se dice, adherirse al externismo supone elegir la teoría más económica, puesto que autoriza la atribución del *mismo* concepto al experto y al profano (relativamente competente en el uso de un término), a pesar de que este último desconozca en parte las condiciones de aplicación del término relevante.²⁹⁴

(3) Mayor estabilidad en el repertorio conceptual de los hablantes

Según lo que acabamos de indicar, el externismo no requiere la postulación de un número potencialmente infinito de conceptos idiosincrásicos en los hablantes. El fenómeno que describimos en (3) constituye una consecuencia de esto, a saber, que de atribuir al hablante profano y al hablante experto el mismo concepto cuando hacen uso del término “T” podremos dotar de mayor estabilidad al repertorio conceptual del hablante profano a lo largo del tiempo, evitando cambios conceptuales inadvertidos como aquellos con los que, según Jackmann (2000, p. 174 y ss.), se ve comprometido el partidario del internismo. Para comprobar este último punto, retomemos el experimento mental de Burge en torno al uso del término “artritis”. El agente que lo protagoniza, quien está dispuesto a asentir ante la preferencia de una serie de oraciones verdaderas

²⁹⁴ No obstante, debe decirse que un internismo sofisticado, que también dé cabida a las relaciones de deferencia lingüística, permite explicar el fenómeno (4). Así, si entre las creencias que un hablante asocia con un término como “oro” se encuentra la creencia de que es aquello a lo que se refieren mediante ese término los hablantes expertos de su comunidad lingüística, entonces el concepto que expresa ese hablante mediante el término “oro” se encuentra determinado, también, por las condiciones de aplicación de ese término establecidas por los expertos de su comunidad.

que involucran el término “artritis” (“tienes artritis desde hace años”, “la artritis es menos dolorosa que el cáncer”, etc.), acude al médico para comunicarle su temor usando la siguiente oración “creo que tengo artritis en el muslo”. Como es bien sabido, la artritis es una dolencia que no afecta a los músculos, dado lo cual el médico corregirá a su paciente y ofrecerá un diagnóstico alternativo. Ante esta situación, como vimos, existen dos lecturas:

- a) El paciente expresa un concepto idiosincrásico, artritis*, distinto del que expresa el médico. Esta lectura es propia de los enfoques internistas: los factores de determinación interna del concepto que expresan el médico y el paciente son distintos y, por tanto, también será distinto el concepto que uno y otro expresan usando el término “artritis”.
- b) El paciente y el médico expresan el mismo concepto, salvo que el paciente tiene una comprensión deficiente del mismo, ya que desconoce que la artritis no afecta a los músculos.

De acuerdo con Jackman, nuestras intuiciones apoyan claramente la lectura externista recogida en (b) porque se trata de una lectura más acorde con el testimonio que probablemente ofrecería el paciente sobre sus propias intenciones lingüísticas. En este sentido, resulta muy verosímil pensar que, tras la conversación con el médico, el paciente corregirá su uso del término “artritis”, reconocerá que se ha equivocado y admitirá, por tanto, que la creencia que expresó al decir “tengo artritis en el muslo” *era una creencia falsa*. Ahora bien, la condición necesaria para que dicha creencia sea falsa es que el concepto que figura en ella sea el concepto de artritis, es decir, el concepto de una enfermedad que no puede ser responsable de los dolores musculares. Dicho de otra manera, si nuestro paciente reconoce que estaba en un error, ello supone que estaba en un error *con respecto a la artritis*, en la medida en que esta dolencia no afecta al muslo. Por tanto, de ser cierto que el concepto que figura en su creencia es el concepto de artritis, entonces nuestro paciente (*antes y después* de la consulta) ha expresado el *mismo* concepto que expresa el médico haciendo uso del término relevante.

En lo referente a la lectura internista descrita en (a), se trata de una opción incompatible con el testimonio del paciente y que, además, atenta contra las condiciones necesarias para que dicho testimonio sea verdadero. La razón es que el partidario de esta tesis mantiene que el concepto que el paciente expresaba mediante el término “artritis” *antes* de hablar con el médico (i) es el concepto idiosincrásico artritis*, que se aplica a una enfermedad que *también* afecta al muslo, y (ii) se trata de un concepto diferente del

concepto que tiene *ahora*, tras aceptar las correcciones pertinentes. Ahora bien, en relación a (i), si este paciente hubiera expuesto sus dudas al médico pensando en la artritis* (en vez de en la artritis), no se habría equivocado, sino que habría estado en lo correcto *con respecto a la peculiar enfermedad a la que se estaba refiriendo*, al ser una dolencia que sí causa dolor en el muslo. Sin embargo, hemos convenido en que el propio paciente es quien reconoce haber errado, por lo que la lectura internista estaría contradiciendo su testimonio. En lo que respecta a (ii), parece claro que el hecho de que nuestro paciente se diga a sí mismo “antes me equivoqué al pensar que la artritis era la causa de mi dolor muscular” manifiesta que, desde su punto de vista, ahora sabe algo más sobre *aquello* (lo mismo) a lo que se refería antes (se diría “cometí un error sobre la artritis, pero ahora sé que *ella* no es la responsable de mi dolor de muslo”). De nuevo, el partidario del internismo se encuentra en una posición desventajosa, puesto que la declaración de nuestro protagonista no evidencia que se haya producido cambio conceptual alguno.

(4) Tomar las palabras en su valor aparente

Algunos autores como Ebbs (1996) consideran que las tesis externistas se adecuan a una máxima implícita que parece regir nuestras prácticas cotidianas de atribución psicológica, a saber, “toma las palabras del hablante en su valor aparente”. Es decir, al observar el comportamiento verbal y no verbal de un sujeto le atribuimos aquellos pensamientos que nos permiten explicar su conducta presente y prever aquella conducta que, según pensamos, es probable que lleve a cabo en el futuro. En este proceso consideramos caritativamente al sujeto de atribución como un hablante competente en el uso de su lenguaje, que pretende maximizar la coherencia lógica entre sus pensamientos y que está dispuesto a revisarlos si advirtiera que entran en conflicto con nuevas evidencias. En caso de que los pensamientos atribuidos fueran inconsistentes o si el comportamiento del agente no coincidiera con las atribuciones realizadas, emprenderemos un proceso de revisión, considerando como primeras opciones que este individuo haya cambiado de opinión, que nos haya mentado, que haya ocultado información, etc. De cualquier manera, el principio de caridad nos insta a tomar como última opción el recurso de declararle irracional y admitir que él mismo permite conscientemente contradicciones entre sus pensamientos o que mantiene pensamientos falsos a sabiendas de que lo son.

Una aplicación del principio de caridad como principio rector de las atribuciones psicológicas consistiría en adscribir al individuo pensamientos cuyo contenido coincide con la proposición expresada por las oraciones que él profiere. La competencia lingüística que presuponemos en el sujeto de atribución nos invita a tomar sus palabras en *su valor aparente*, es decir, como si este hablante las usara para expresar los conceptos que *se* expresan en su lenguaje mediante el uso de dichas palabras. Salvo que nuestros intercambios lingüísticos con este agente nos proporcionen evidencias de lo contrario, la interpretación que prevalecerá será la interpretación de sus preferencias en su valor aparente.

El fenómeno descrito sintoniza bien con la tolerancia que el externismo social y mixto muestra ante el fenómeno del dominio incompleto de los conceptos. De este modo, la determinación social o, al menos, parcialmente social del contenido permite tomar a los agentes en tanto que hablantes que, salvo evidencia de lo contrario, usan sus términos como *se* usan (en su lenguaje).

2.3 *Tensiones internas a la clasificación usual de las variedades del externismo*

De acuerdo con lo expuesto, comenzaremos la tercera y última parte de esta sección examinando la reconstrucción que ofrece Burge del experimento mental de la Tierra Gemela en torno al uso del término “agua”, con el fin de precisar de qué manera este autor ha conseguido equilibrar la influencia de los factores naturales y sociales en la individuación de nuestros contenidos. Este análisis nos servirá para poner de manifiesto que las tesis del externismo mixto o amplio (que es la comprensión externista a la que Burge dice adherirse) se ven lastradas por tensiones internas, especialmente en lo relativo a la individuación de los conceptos expresados por términos de género natural. En este sentido, argumentaremos que existen ciertas situaciones ante las cuales los elementos naturales y los elementos sociales fijan una identidad distinta para los conceptos relevantes, por lo que desde una posición externista amplia se deberán hacer esfuerzos adicionales para decidir a qué clase de factores (naturales o sociales) se desea otorgar preeminencia.

En relación a las ideas de Burge, vimos que este autor describe su opción externista, que él denomina “*anti-individualismo*”, como una tesis que recoge la influencia de los factores del entorno externo al hablante en la determinación de sus contenidos mentales, especialmente aquellos en cuya expresión lingüística figuran

términos generales.²⁹⁵ Así, aunque el término que figura en su experimento mental más famoso (“arthritis”) no es, para este autor, un término de género natural,²⁹⁶ Burge considera que su posición como anti-individualista *también* da cabida a la influencia del entorno físico en la determinación de los contenidos mentales de los hablantes (Burge 1982, p. 84). Por este motivo, parece que su teoría debería comprenderse dentro del externismo amplio o mixto. No obstante, antes de examinar si esta interpretación es la más adecuada, conviene reparar en la siguiente cuestión. De acuerdo con la caracterización que hemos ofrecido de las distintas variedades de externismo, tanto el externismo social como el externismo mixto son tesis sobre la individuación de los conceptos expresados por términos generales, entre los que se encuentran los términos de género natural. El externismo natural, por su parte, propone una individuación externista de los conceptos expresados únicamente por nuestros términos de género natural. En cuanto al externismo mixto, sus valedores sostienen, por un lado, que los factores naturales no tienen influencia individuadora en lo que respecta a los conceptos expresados por términos generales distintos de los términos de género natural, pero sí tienen una influencia (parcial) sobre la individuación de los conceptos expresados por términos de género natural.²⁹⁷ Por otro lado, los factores sociales tienen una influencia individuadora (parcial) en relación a los conceptos expresados *tanto* por términos de género natural *como* por términos generales distintos de los de género natural. La consecuencia de ello es que el externismo mixto ha de conjugar la influencia de los factores sociales y naturales *sólo en relación a los conceptos expresados por los términos de género natural*. Por esta razón, para descubrir en cuál de las concepciones del externismo resulta más conveniente situar el anti-individualismo de Burge, debemos acercarnos a su trabajo de (1982), dedicado a extrapolar al terreno conceptual algunas

²⁹⁵ Burge señala que el experimento mental planteado en torno al uso del término “arthritis” podría haberse construido también usando “un término de artefacto, una palabra ordinaria de género natural, un adjetivo de color, un término de rol social, un término de un estilo histórico, un nombre abstracto, un verbo de acción, un movimiento físico” (Burge 1979, pp. 106-107). Aunque este autor prefiere dejar abierto el alcance de las conclusiones anti-individualistas, parece que todos los términos (generales) que designan “clases públicas de objetos, propiedades y eventos” son susceptibles de recibir una individuación externista (Burge 1986c, p. 223).

²⁹⁶ Sin embargo, esta posición no está exenta de controversia. Putnam, por ejemplo, considera que la esclerosis múltiple es un género natural, tal y como lo son los géneros naturales caballo y oro (Putnam 1983b, p. 71).

²⁹⁷ Las tesis de Putnam (1975c) sobre la referencia de los términos de género artificial, que sí reconocen una naturaleza o esencia en los ejemplares que pertenecen a esta clase de géneros como factor determinante para la aplicación del término, no han gozado de tanto éxito como su teoría de la referencia de los términos de género natural. En particular, no conocemos ningún partidario del externismo psicológico que haya intentado usar estas tesis en relación a la individuación de conceptos de género artificial.

de las conclusiones más importantes que Putnam expuso en (1975c) sobre la semántica de los términos de género natural.

A este respecto Burge afirma expresamente que la visión de Putnam “complementa y suplementa” su propia posición (Burge 1979, p. 108, n. 2). En particular, este autor indica que acepta las tesis de Putnam sobre la semántica de los términos de género natural en la medida en que nos llevan a concluir que dos hablantes internamente idénticos que usan un mismo término de género natural para referirse a dos géneros naturales cuyas propiedades subyacentes son distintas pero que, al mismo tiempo, resultan macroscópicamente indistinguibles, son agentes que expresan, cada uno de ellos, un concepto distinto, al tener el término relevante un significado diferente en cada caso. Ahora bien, esta idea no coincide exactamente con la conclusión de Putnam en (1975c) ya que, como se recordará, en ese trabajo este autor postulaba una identidad en los estados mentales (en sentido restringido) de Óscar₁ y Óscar₂. Si consideramos que los conceptos son los constituyentes de los contenidos de esos estados mentales (en sentido restringido), entonces el postulado de Putnam tendrá como consecuencia que los conceptos que estos sujetos expresan usando el término “agua” serán exactamente idénticos. Por este motivo, aunque las ideas expuestas en Putnam (1975c) nos permiten concluir que una diferencia en la estructura subyacente de los géneros naturales determina una diferencia en el *significado* de un término de género natural que usan dos hablantes internamente idénticos, ya mostramos en el primer capítulo que por sí mismas no permiten establecer también una diferencia en el *concepto* expresado por dicho término. Para dar este último paso, Burge dedica considerables esfuerzos a combatir aquel aspecto de las tesis de Putnam que, a su juicio, es responsable de la resistencia a establecer este tipo de diferencias conceptuales, a saber, la idea de que “los términos de género natural son indéxicos” (Burge 1982, p. 87).²⁹⁸ Puesto que en la tercera sección del primer capítulo del presente trabajo ya analizamos los argumentos esgrimidos por Burge para combatir la concepción indéxica de los términos de género natural, no nos extenderemos más sobre esta cuestión. Lo más relevante a este respecto es que para Burge no tiene sentido establecer una distinción entre estados mentales en sentido estrecho (internamente determinados) y estados

²⁹⁸ Ya se ha indicado que ésta es una descripción excesivamente simple (cuando no falsa) de las tesis de Putnam sobre los términos de género natural enunciadas en (1975c). Recordemos que Putnam no considera que los términos de género natural sean indéxicos, sino que existe un componente indéxico en las definiciones que suelen usarse para fijar la referencia de esta clase de términos. Además de (1982), Burge discute la visión indéxica de los términos de género natural en (1979) p. 108, n. 2, (1995) p. 341, n. 5, (1986c) p. 267, n. 12, (2006) p. 152 y (2007) p. 11.

mentales en sentido amplio, dado lo cual los aspectos intrínsecos de los protagonistas del experimento (idénticos por hipótesis) en relación al líquido que denominan “agua” no servirán para determinar un genuino contenido mental (idéntico en ambos sujetos).

Ahora bien, aunque Burge dice aceptar en lo esencial las conclusiones de Putnam para dar cabida a la determinación de contenidos en virtud de factores externos de carácter exclusivamente natural, la clasificación de este autor en tanto que partidario del externismo mixto requiere algunas matizaciones. La mejor manera de comprobarlo es atender a la reconstrucción que Burge nos ofrece en (1982) del experimento mental de Putnam en torno al uso del término “agua”. Como se ha indicado, éste es el único experimento perteneciente al texto de Putnam de (1975c) que respalda una versión estrictamente natural del externismo, ya que se plantea en relación a un término con respecto al cual no hay división del trabajo lingüístico en la comunidad de hablantes.

Lo primero que debe decirse acerca de la reconstrucción de Burge de dicho experimento es que en ella se usan exactamente las mismas premisas empleadas por Putnam *pero con dos modificaciones*, a saber, que Óscar₁ y Óscar₂ no comparten un estado mental en sentido restringido y que en las respectivas comunidades científicas de estos sujetos un grupo de hablantes conoce la estructura química de los líquidos en cuestión (Burge 1982, p. 85). El motivo para efectuar la primera modificación parece claro, puesto que si se pretende usar el experimento mental de Putnam para extraer conclusiones sobre la individuación de los estados mentales de sus protagonistas, no parece legítimo postular de partida una identidad o diferencia en dichos estados. Pero, por otro lado, la justificación de Burge para variar el segundo aspecto señalado resulta un tanto sorprendente. Según nos dice este autor, él supone que en las comunidades lingüísticas de Óscar₁ y Óscar₂ hay conocimiento de la estructura subyacente de los líquidos porque, aunque se trata de una “extensión significativa de las ideas de Putnam [...] no es necesario para [sus] propósitos primarios” postular la ignorancia social de esa propiedad (Burge 1982, p. 86, n. 8). Sin embargo, a nuestro modo de ver, conservar este supuesto es necesario para demostrar que no sólo los usos comunitarios de las palabras, sino también los *factores del entorno físico influyen por sí mismos* en la determinación del contenido psicológico, lo cual, según vimos, constituye el propósito primario de Burge en (1982). Desde el instante en que se incorpora al escenario de Putnam el conocimiento de la microestructura a los usos socialmente determinados – aun cuando esto sólo ocurre con respecto a una pequeña parte de la comunidad lingüística –, este experimento modificado no sirve para extraer conclusiones sobre el poder de

individuación que, de suyo, poseerían presuntamente los factores naturales, puesto que, si tienen alguna influencia en este sentido, no podemos saber si se trata de una influencia directa o de una influencia mediada por los usos sociales de los términos relevantes.

Podemos explicar esto de otra manera. Al agregar al escenario original de Putnam una diferencia en los usos comunitariamente establecidos del término “agua” no es posible apreciar si el factor responsable de la diferencia en los conceptos que expresan los sujetos de cada entorno mediante el uso del término “agua” es o bien la diferencia en los usos lingüísticos asociados al término, o bien la diferencia en la estructura subyacente de los líquidos relevantes. Por este motivo, la segunda modificación que Burge efectúa con respecto al experimento original de Putnam no permite extraer conclusiones favorables al externismo natural. No obstante, pensamos que dicha modificación no es en absoluto un descuido, sino una señal de que la tesis que este autor está defendiendo no es un externismo mixto tal y como lo hemos descrito sino, en último término, una tesis más cercana al externismo social. En este sentido, puede observarse que en las páginas que siguen a la reconstrucción del experimento mental de Putnam llevada a cabo en (1982), Burge destaca en cursivas el hecho de que la palabra “agua” *interpretada en castellano* no tiene el mismo significado que ese término *interpretado en castellano-gemelo* (1982, pp. 88-89). Esto es, el pivote de la argumentación anti-individualista de Burge reside no ya en las diferentes propiedades subyacentes de los líquidos (como en el caso de Putnam), sino en el hecho de que el significado comunitario del término “agua” no es el mismo en la comunidad lingüística de los castellanohablantes y en la comunidad lingüística de los castellano-hablantes-gemelos. Por este motivo, una preferencia de la oración “el agua es húmeda” por parte de un habitante de la Tierra Gemela no expresará el pensamiento de que el agua es húmeda, ya que, al estar inserto en una comunidad de hablantes que asigna un significado distinto al término “agua”, no es posible tomar sus usos del término en su valor aparente para nosotros, los castellanohablantes. Como puede apreciarse, esta clase de razonamiento es impecable desde el punto de vista del externismo estrictamente social.

Pese a esto, Burge considera que la interpretación que ha ofrecido del escenario de la Tierra Gemela permite mostrar cómo el anti-individualismo *incorpora* las conclusiones fundamentales de Putnam en el terreno de la determinación de nuestros contenidos mentales. Según afirma, en tanto que los líquidos relevantes tienen una

estructura subyacente distinta, podremos establecer una diferencia en el concepto que Óscar₁ y Óscar₂ expresan mediante el uso del término “agua”. Esta tesis de Burge revela que comparte con Putnam cierto aspecto de su teoría de la referencia de los términos de género natural de (1975c), a saber, que la extensión de este tipo de términos viene determinada por la posesión de una estructura subyacente o esencia (la composición química en el caso de las sustancias como el agua). Por este motivo, tanto para Putnam como para Burge el género natural agua está individuado por su composición química, dado lo cual una porción de líquido será agua (y debería recibir la denominación de “agua”) sólo si comparte su composición química con los miembros paradigmáticos del género (Burge 2006 p. 168). El hecho de que Burge se adhiera de este modo a la teoría de la referencia de los términos de género natural de Putnam, que tiene como consecuencia que sólo las porciones de H₂O pertenecen a la extensión del término castellano-terrestre “agua”, sustenta la concepción de que los pensamientos que Óscar₁ y Óscar₂ expresan usando el término “agua” son pensamientos con contenidos distintos. Estos hablantes no piensan lo mismo cuando dicen “el agua es húmeda” por la sencilla razón de que uno piensa en el líquido que tiene como composición química H₂O y el otro piensa en el líquido que tiene como composición química XYZ. En consecuencia, dice Burge, no sólo los factores del entorno socio-lingüístico sino *también* los factores del entorno físico (la estructura subyacente de los géneros naturales) ejercen una influencia individuadora.

Sin embargo, reparemos en la siguiente afirmación de Burge: “[L]a diferencia en los estados mentales [de Óscar₁ y Óscar₂] parece ser, en primera instancia, un producto de diferencias en su entorno físico, mediadas por diferencias en sus entornos sociales – en los estados mentales de sus semejantes y en el significado convencional de las palabras que emplean ellos y sus semejantes” (Burge 1982, p. 87). En efecto, es porque existe una diferencia en la estructura subyacente de los líquidos, que determina una identidad distinta para cada uno de ellos, por lo que establecemos una diferencia en el concepto que Óscar₁ y Óscar₂ expresan haciendo uso del término “agua”. Sin embargo, Burge deja muy claro que esta diferencia conceptual está *mediada* por una diferencia en el uso socialmente establecido para dicho término. En otras palabras, el conocimiento de las propiedades subyacentes está incorporado al patrimonio de conocimiento social asociado al término, algo que no era en absoluto necesario para Putnam. Desde luego, aunque esto no significa que para Burge todo hablante competente en el uso del término “agua” debe conocer (tanto en la Tierra como en la Tierra Gemela) cuál es la estructura

subyacente del género relevante, este autor reconstruye el experimento mental añadiendo el hecho de que en las comunidades científicas de ambos entornos se tiene noticia de la diferente estructura subyacente de lo que en ambos casos se llama “agua”. Por este motivo, pensamos que el anti-individualismo de Burge debe recibir una caracterización más sofisticada que la recogida en nuestra descripción del externismo mixto, una teoría que parece reservar un grado de influencia equivalente a los factores naturales y a los factores sociales. Así, en la medida en que para este autor la influencia de los factores naturales en la individuación de aquellos conceptos expresados por términos de género natural *está mediada por* la influencia de los factores sociales a este respecto, lo realmente determinante para establecer la identidad de tales conceptos es esta última clase de factores. Para Burge, por tanto, cabe otorgar un papel a la influencia de los elementos externos naturales (propiedades subyacentes de los géneros) a la hora de fijar la identidad de nuestros conceptos de género natural, pero se trata de una *influencia secundaria, mediada por y dependiente de* la influencia que ejercen los factores sociales.

Dicho todo esto, proponemos la siguiente distinción que constituirá el primer paso para diseñar una tipología más compleja entre las variantes de externismo. Por un lado, distingamos la posición de Burge, que establece una jerarquía entre los factores naturales y los factores sociales, de tal manera que sólo existiría una diferencia conceptual si la diferencia entre los géneros naturales relevantes quedara recogida en el cuerpo de conocimiento social asociado al término en cuestión. Por otro lado, también cabría la posibilidad de tomar la opción contraria, en cuyo caso una diferencia en los géneros naturales supondría una diferencia en el concepto expresado por un mismo término, aun cuando esta diferencia no se viera reflejada en los usos sociales del término. En este último caso, se estaría abogando por una individuación natural directa, tal y como la hemos definido. Dadas estas dos opciones, cualquier posición externista mixta, esto es, cualquier posición que reconozca, además de la influencia de los factores naturales, la influencia de los factores sociales en la individuación de los conceptos expresados por términos generales, incluidos los términos de género natural, podrá posicionarse de dos formas diferentes en relación a los factores que determinan qué conceptos expresamos haciendo uso de esta última clase de términos:

- (a) Asumir las tesis del externismo natural y mantener, en consecuencia, que la identidad de los conceptos expresados mediante términos de género natural depende en parte pero fundamentalmente de las propiedades subyacentes de los ejemplares que pertenecen a los géneros naturales, aun cuando el conocimiento de dichas propiedades no haya quedado incorporado a los usos comunitarios del término de género natural relevante.
- (b) La identidad de los conceptos expresados mediante términos de género natural depende en parte pero fundamentalmente de las propiedades subyacentes de los ejemplares que pertenecen a los géneros naturales, siempre y cuando el conocimiento de esas supuestas propiedades haya quedado incorporado a los usos comunitarios del término de género natural relevante.

A partir de esto, podemos comprobar que un partidario del externismo mixto que opte por la tesis (a) *distribuirá* la influencia de los factores naturales y sociales del siguiente modo. Por un lado, los factores naturales concebidos como propiedades subyacentes tendrán influencia parcial en la individuación de los conceptos expresados por términos de género natural y, por otro lado, los factores sociales tendrán influencia parcial en la determinación de los conceptos expresados por términos generales distintos de los términos de género natural. Por su parte, quienquiera que abogue por un externismo mixto combinado con la tesis (b) estará defendiendo en último término una variante del externismo social, a saber, aquella que *requiere que el conocimiento de las propiedades subyacentes de los géneros naturales pertenezca al dominio social* para que dichas propiedades ejerzan una influencia en la individuación de los conceptos relevantes.²⁹⁹ Hecha esta aclaración, parece que las posturas externistas inspiradas en el experimento mental de Putnam en torno al término “agua” de (1975c) pueden identificarse con un externismo mixto complementado con la tesis (a), ya que para los términos de género natural esta opción acepta las tesis del externismo natural, mientras que las propuestas de Burge encajan mejor en un externismo mixto complementado con la tesis (b).

En este punto sería recomendable preguntarse por qué un partidario del externismo mixto debería tomar los complementos (a) y (b) como opciones excluyentes.

²⁹⁹ Según comprobaremos en la próxima sección, ésta no es la única variante posible del externismo social, en la medida en que cabría proponer una concepción de esta tesis que asigne un poder individuador a los factores naturales – socialmente mediados – aun cuando dichos factores no fueran propiedades subyacentes (i.e., determinantes esenciales de la extensión de los géneros naturales).

Responderemos a esta cuestión basándonos en el examen de una serie de casos en los cuales la combinación del externismo mixto con la tesis (a), por un lado, y la combinación del externismo mixto complementado con la tesis (b), por otro, determinarán en cada caso una identidad distinta para el concepto expresado por un término de género natural.

El primer caso concierne al concepto expresado por el término “agua” en el escenario descrito por Putnam en (1975c) que, como vimos, es distinto al que Burge reconstruye en (1982). Según vimos, el propósito de Putnam al plantear ese experimento mental es aportar plausibilidad a la tesis de que “las convenciones sociales/lingüísticas no pueden determinar completamente el significado” (Liu 2002, pp. 286 y ss.). Es decir, que aun cuando las reglas de uso que ambas comunidades establecen para el término “agua” son las mismas, el significado del término difiere en un entorno y en otro. Ahora bien, si no existe ninguna diferencia en las convenciones lingüísticas que rigen el empleo del término “agua” en las respectivas comunidades de Óscar₁ y Óscar₂, tampoco existiría ninguna diferencia en el concepto expresado por estos hablantes mediante el uso de ese término si respaldáramos la opción mixta combinada con la tesis (b).

Liu indica, además, un segundo caso en el cual ambas concepciones del externismo mixto entran en tensión, a saber, los ejemplos del “error de los expertos” (2002, p. 390). Supongamos que en 1790 nuestra comunidad científica hubiera determinado (erróneamente) que la composición química del agua es XYZ y supongamos, además, que nadie hubiera remediado jamás esta equivocación. Dada esta situación, de acuerdo con el externismo mixto en combinación con (a), el concepto que nosotros expresaríamos usando el término “agua” sería distinto del concepto que expresarían los habitantes de la Tierra Gemela haciendo uso del mismo término. Esta diferencia existiría aun cuando nosotros creyéramos que es uno y el mismo concepto, basándonos en la suposición (errónea) de que ambos tipos de líquido tienen exactamente la misma composición química. En cambio, el externismo mixto conjugado con la tesis (b) ofrecerá la respuesta contraria, a saber, que Óscar₁ y Óscar₂ expresarán el mismo concepto haciendo uso del término “agua”, en cuanto que los usos asociados a ese término en las comunidades de la Tierra y de la Tierra Gemela son

exactamente los mismos (se aplica al líquido inodoro, incoloro e insípido cuya composición química se ha determinado como XYZ).³⁰⁰

El tercer tipo de casos concierne a las situaciones de traslado inadvertido que tuvimos ocasión de examinar en detalle en el tercer capítulo de esta investigación. Según comprobamos, los externistas partidarios de la tesis (b) no tienen por qué verse comprometidos con la tesis de que Óscar₁, al ser transportado inadvertidamente a la Tierra Gemela, termina por expresar un concepto distinto del que expresaba en la Tierra mediante el uso del término “agua”. En tanto que las convenciones socialmente establecidas para el uso de ese término son exactamente las mismas en la Tierra y en la Tierra Gemela esta teoría no tendría motivos para establecer una diferencia en el concepto expresado por los usos del término en un entorno y otro.

Estos tres ejemplos nos inducen a pensar que un partidario del externismo mixto ha de tomar como opciones excluyentes combinar su postura relativa a la individuación de los conceptos expresados por los términos de género natural o bien con la tesis (a), o bien hacerlo con la tesis (b). La razón es que, de no establecer una jerarquía en la influencia de los factores naturales, por un lado, y sociales, por otro, surgirán conflictos en los casos siguientes.³⁰¹ En primer lugar, cuando existan diferencias en las propiedades subyacentes de los géneros naturales que no estén registradas en los usos sociales de los términos de género natural, la individuación en función de esta clase de propiedades de los géneros naturales será más fina que la individuación en función de factores sociales. Ésta es la situación que se da, precisamente, en relación al término “agua”, un término que designa dos géneros naturales distintos en la Tierra y en la Tierra Gemela según la concepción de los géneros naturales de Putnam, pero que, al mismo tiempo, viene regido por las mismas convenciones lingüísticas. En segundo lugar, podría ocurrir lo contrario, es decir, que los criterios sociales den lugar a una individuación más fina que la individuación en función de factores estrictamente naturales. Esto ocurriría, por ejemplo, si las normas sociales regulan el uso del término de género natural “T” de tal manera que se sanciona como correcta su aplicación al género en cuestión pero sólo bajo cierto estado, mientras que otro término distinto, “T₁” es el adecuado para denominar al género relevante bajo un estado diferente. Es digno de señalarse que al parecer este fenómeno se produjo, precisamente, en relación al término

³⁰⁰ Cf. Beattie (1993), p. 319.

³⁰¹ Los tres ejemplos examinados se agruparían dentro del primer tipo de casos, a saber, cuando la influencia natural directa da como resultado una individuación más fina que la individuación en función de los usos sociales.

“agua” en 1750, la época en que Putnam sitúa el escenario de la Tierra Gemela: las convenciones que regulaban su uso restringían su aplicación al agua en estado líquido (Kuhn 1989, p. 28).

De lo dicho hasta el momento se desprende que existen ciertas tensiones en el seno de un externismo mixto en lo que se refiere al grado de influencia que debe asignarse a los factores naturales y a los factores sociales en la individuación de los conceptos de género natural. La existencia de esta clase de conflictos es algo que ha pasado casi desapercibido en la literatura especializada.³⁰² Según pensamos, cabe explicar esta considerable falta de atención mediante la siguiente hipótesis. El lector de los experimentos mentales, el tercero que juzga si ha de establecerse una diferencia en los conceptos que dos hablantes expresan mediante el uso de un término, es un hablante que vive en un entorno (espacio/temporal) en el que, desde su punto de vista, los factores sociales y los factores naturales desempeñan su rol individuador de manera acompañada. El planteamiento de un experimento mental no se lleva a cabo *desde el punto de vista del Ojo de Dios*, sino desde el punto de vista de ciertos hablantes (autor y lector) insertos en una comunidad de hablantes que determina ciertos usos para los términos de género natural en función del conocimiento que se tiene en esa misma comunidad de la naturaleza de estos géneros. Para nosotros, lectores de Putnam, que sabemos que existe una diferencia entre los líquidos relevantes de la Tierra y de la Tierra Gemela, es relativamente intuitivo pensar que el término “agua” tiene un significado distinto en un entorno y otro. Desde nuestro punto de vista, las tensiones entre los factores sociales y los factores naturales tienden a reducirse al mínimo por la sencilla razón de que nosotros mismos aplicamos los términos de género natural en función de lo que sabemos (sobre los géneros naturales) y, al revés, nuestro conocimiento de los géneros naturales se recoge usando una terminología cuyo uso está sancionado socialmente en nuestro entorno.

Antes de comenzar la tercera sección, en la que propondremos una nueva clasificación de las concepciones del externismo, conviene recapitular los aspectos más importantes de lo dicho hasta el momento. En primer lugar, ofrecemos una clasificación de las distintas variantes de externismo, así como una enumeración de los fenómenos relacionados con el lenguaje y el pensamiento en los que estas tesis encuentran apoyo. En segundo lugar, comprobamos que resulta bastante cuestionable que el anti-

³⁰² Constituyen excepciones Beattie (1993), Rudd (1997), Marconi (1997), Esfeld (2002) y Liu (2002).

individualismo defendido por Burge pueda quedar bien caracterizado como un externismo mixto sin cualificaciones. En este sentido, hemos argumentado que este autor no asigna el mismo grado de influencia a los factores naturales y a los factores sociales, sino que son estos últimos los que someten la influencia de los primeros. Finalmente, tratamos de explicar por qué es necesario introducir ciertas modificaciones en el seno del externismo mixto. Para ello, examinamos una serie de situaciones que ponen en evidencia las profundas tensiones que existen entre la individuación de los conceptos de género natural en función de factores naturales y la individuación de estos conceptos en función de factores sociales.

3. *Propuesta de revisión de la tipología usual del externismo*

De acuerdo con lo dicho en las páginas precedentes, hay motivos para pensar que la clasificación de las distintas variantes de externismo psicológico que suele ofrecerse en la literatura al uso y que nosotros recogimos en la sección anterior presenta algunas deficiencias. Nuestra tarea en esta sección consiste en proponer una nueva clasificación que evite estos problemas.

Según comprobamos, la dificultad fundamental que presenta la clasificación ofrecida en la sección anterior consiste en una cierta indeterminación en el seno de las tesis del externismo mixto. Hemos visto que existen situaciones en las cuales los factores naturales y los factores sociales determinan, en cada caso, una identidad distinta para el concepto de género natural relevante. Por este motivo, enunciaremos dos posibles complementos para la posición mixta, en función de si son los factores naturales o los factores sociales los que gozan de una influencia determinante en relación a los conceptos expresados por términos de género natural.³⁰³ De acuerdo con esto, en caso

³⁰³ Debe realizarse una aclaración en lo que respecta al uso de las expresiones “influencia determinante” o “predominante” de un tipo de factores sobre otros. Al afirmar que un partidario del externismo mixto complementado con la tesis (a) otorga una influencia determinante a los factores naturales con respecto a la individuación de los conceptos expresados por términos de género natural lo que queremos indicar es que dicha influencia *predomina sobre la influencia de los factores sociales*. Para esta concepción lo *determinante* a la hora de establecer una diferencia en los conceptos expresados por dos hablantes que habitan en entornos distintos cuando hacen uso de un término de género natural es que exista una diferencia en los factores naturales (propiedades subyacentes de los géneros naturales) de un entorno y otro, al margen de que los factores sociales (las convenciones asociadas al término en cuestión) sean los mismos (o no) en ambos entornos. En el caso del externismo mixto complementado con la tesis (b), la influencia de los factores sociales es *determinante con respecto a la influencia de los factores naturales* en el sentido de que sólo existirá una diferencia en el concepto expresado por dos hablantes pertenecientes a entornos distintos mediante el uso de un término de género natural si las convenciones que en un entorno y otro se asocian con el término en cuestión son diferentes. Declarar que la influencia

de conflicto entre estos dos factores un partidario del externismo mixto podría otorgar una influencia predominante a las propiedades subyacentes que determinan los géneros naturales. Por este motivo, si el conocimiento de dichas propiedades subyacentes no está incorporado a las normas públicas que rigen el uso del término, ello no obsta para que dichas propiedades ejerzan una influencia individuadora directa sobre nuestros conceptos de género natural. En cambio, la opción a la que Burge parece adherirse requiere que las diferencias en las propiedades subyacentes de los géneros naturales estén incorporadas a los usos socialmente establecidos de los términos de género natural para que tengan un papel en la individuación de contenidos. Por ello, para un partidario del externismo mixto en combinación con la tesis (b) los factores que determinan la identidad de los conceptos de género natural han de estar mediados, en última instancia, por el conocimiento social.

Es conveniente preguntarse si las tesis del externismo mixto conjugadas con (b) constituyen la única manera de comprender el externismo social. En nuestra opinión, esta cuestión merece una respuesta negativa. Como se ha indicado, la caracterización del externismo social ofrecida en la sección anterior no alude en ningún momento a las propiedades subyacentes de los géneros naturales. El externismo social es, en principio, independiente de la tesis que afirma que los géneros naturales están individuados por sus propiedades subyacentes. Por este motivo, cabría la posibilidad de defender una variante de externismo social que fuera combinada con otra concepción de los géneros naturales, sin distinguir entre los criterios socialmente establecidos para la aplicación de un término de género natural una propiedad o un conjunto de propiedades en tanto que determinantes fundamentales del género.

Por estas razones, proponemos una nueva clasificación de las variedades del externismo sobre la individuación de nuestros pensamientos. La primera variante se corresponderá con la clase de externismo que reconoce la individuación natural directa con respecto a los conceptos expresados por los términos de género natural y no se pronuncia en relación a la posible influencia de los factores externos (de carácter social) en la individuación de conceptos expresados por otra clase de términos generales. Por otra parte, la bifurcación entre el externismo mixto conjugado con (a) y el externismo

de un factor externo (natural o social) es determinante o predominante con respecto la influencia de otro factor externo (natural o social) no equivale en ningún caso a la afirmación de que para establecer la identidad de un concepto *debe atenderse exclusivamente a factores de carácter externo, ya sean naturales o sociales*. Para todas las variantes de externismo caracterizadas, la influencia de los factores externos (naturales o sociales) siempre es parcial con respecto a la influencia de los factores internos.

mixto conjugado con (b) nos obliga a diferenciar entre dos tipos de externismo mixto, a saber, un externismo de distribución (correspondiente al primer miembro de esta bifurcación) y un externismo estrictamente social (correspondiente al segundo miembro). En último lugar, queremos introducir una distinción adicional en el seno del externismo estrictamente social por lo que se refiere a la individuación de los conceptos expresados por términos de género natural para dar cabida a una posible concepción no esencialista de esta tesis:

Externismo estrictamente natural: La identidad de los contenidos mentales que un hablante expresa usando *términos de género natural* depende, en parte pero fundamentalmente, de las propiedades subyacentes de los miembros de la extensión de tales términos, concebidas según la teoría de los géneros naturales de Putnam. Esta tesis se concreta de la siguiente manera: dos hablantes exactamente idénticos en todos sus aspectos internos pero pertenecientes a entornos distintos expresarán conceptos diferentes al usar el término de género natural “T” si los ejemplares que pertenecen a la extensión de “T” poseen, en sus respectivos entornos, una estructura subyacente diferente. Esto es así incluso aunque los ejemplares que reciben el nombre de “T” en cada entorno sean indistinguibles para cualquier hablante de cualquiera de estos entornos (i.e. aunque nadie conozca esa diferencia en las propiedades subyacentes de los géneros).

Externismo de distribución: La identidad de los contenidos mentales que un hablante expresa usando un *término general (excepto si se trata de un término de género natural)* depende, en parte pero fundamentalmente, del uso que la comunidad lingüística de dicho hablante ha establecido para dicho término. Esta tesis se concreta de la siguiente manera: dos hablantes exactamente idénticos en todos sus aspectos internos que están insertos, cada uno de ellos, en comunidades lingüísticas que establecen convenciones diferentes para el uso del término “T” expresarán conceptos distintos mediante el uso de dicho término. Esto es así aunque tales sujetos desconozcan la diferencia en las convenciones lingüísticas comunitarias que rigen el uso del término.

Por lo que respecta a la identidad de los contenidos mentales que un hablante expresa usando un *término de género natural*, ésta viene determinada según las pautas del externismo estrictamente natural.

Externismo social: La identidad de los contenidos mentales que un hablante expresa usando un *término general (de género natural o no)* depende, en parte pero fundamentalmente, del uso que la comunidad lingüística en la que está inserto dicho hablante ha establecido para dicho término. Esta tesis se concreta de la siguiente manera: dos hablantes exactamente idénticos en todos sus aspectos internos que están insertos, cada uno de ellos, en comunidades lingüísticas que establecen convenciones diferentes para el uso del término “T” expresarán conceptos distintos mediante el uso de dicho término. Esto es así incluso aunque tales sujetos (en particular) desconozcan la diferencia en las convenciones lingüísticas comunitarias que rigen el uso del término.

En el caso de los términos de género natural, la relación entre las convenciones sociales (determinantes en última instancia) y los factores naturales puede concebirse de dos maneras distintas:

Externismo social esencialista: Los conceptos expresados mediante términos de género natural son individuados en función de las propiedades subyacentes de los ejemplares que pertenecen a los géneros naturales, concebidas según la teoría de los géneros naturales de Putnam, siempre y cuando el conocimiento de dichas propiedades haya quedado incorporado a los usos comunitarios del término de género natural relevante.

Externismo social no-esencialista: Los conceptos expresados mediante términos de género natural son individuados en función de los usos comunitariamente asignados a estos términos sin necesidad de que estos usos apelen a propiedades subyacentes caracterizadas según la teoría de los géneros naturales de Putnam aunque tales usos sí pueden hacer mención a una o (generalmente) varias propiedades naturales que no tienen por qué cumplir el rol de las propiedades subyacentes – siendo el criterio de pertenencia a un género la posesión de la mayoría de estas propiedades (no de todas ellas) –. Estas propiedades naturales pueden tener un carácter funcional, o, incluso, podrá tratarse de propiedades que

según la teoría de Putnam son superficiales, pero que en los usos ordinarios de un término de género natural tienen una importancia destacada a la hora de otorgar capacidades discriminativas a los hablantes con respecto al uso de otros términos de género natural.

Una vez establecida esta nueva clasificación de las variedades de externismo, conviene realizar las siguientes observaciones. En primer lugar, el externismo estrictamente natural no se pronuncia sobre las condiciones de individuación de los conceptos expresados por los términos generales distintos de los de género natural. Debido a ello, esta teoría podría ser respaldada por alguien que tuviese una opinión cercana al internismo en lo que se refiere a esta última cuestión. En segundo lugar, en lo que se refiere al externismo de distribución, esta posición distribuye la influencia de los factores naturales y los factores sociales en la individuación de nuestros conceptos de la siguiente manera. Por un lado, se mantiene que los usos comunitariamente establecidos para nuestros términos contribuyen a determinar la identidad de los conceptos expresados por los términos generales distintos de los términos de género natural, mientras que los factores naturales – las propiedades subyacentes de los géneros – intervendrían únicamente en la individuación de los conceptos expresados por los términos de género natural. Esta posición tiene la ventaja de evitar las interferencias entre el poder individuador de los factores sociales, por un lado, y de los factores naturales, por otro. En tercer lugar y con respecto al externismo social, hemos optado por ofrecer una subdivisión en su seno atendiendo a dos posibles concepciones de los factores naturales (incorporados a los usos lingüísticos de la comunidad) que determinan la identidad de los conceptos. La primera de estas subdivisiones distinguiría un externismo social esencialista, que considera que los factores naturales (incorporados al conocimiento social) que contribuyen a la determinación de los conceptos que expresamos haciendo uso de términos de género natural quedan constituidos por las propiedades subyacentes de los géneros naturales, de acuerdo con la teoría que mantienen al respecto Putnam y Kripke. Podemos advertir que el anti-individualismo defendido por Burge encaja en esta variante, ya que tanto para Putnam como para Burge la extensión de esta clase de términos viene determinada por la posesión de una(s) propiedad(es) subyacente(s), revelada por la investigación científica (Burge 2006 p. 168). De este modo, la diferencia en las propiedades subyacentes de los líquidos que en la Tierra y en la Tierra Gemela se denominan “agua”, propiedades que han quedado

recogidas en los usos lingüísticos asociados al término en ambas comunidades, conlleva que el concepto que expresan Óscar₁ y Óscar₂ usando el término “agua” sea un concepto diferente para cada uno de ellos. No obstante, debemos reservar otra subdivisión para el externismo social no-esencialista, según el cual la individuación de los conceptos expresados por los términos de género natural depende de los usos socialmente establecidos para esta clase de términos *sin requerir* que dichos usos se establezcan a partir del conocimiento de presuntas propiedades subyacentes que determinan la extensión de los géneros. Esto no significa que para esta comprensión del externismo no existan algunas propiedades especialmente relevantes para la constitución de un género natural y, por tanto, para la determinación del concepto correspondiente. Sólo se rechaza que haya de existir, para cada género natural, el tipo de propiedades que cumplen el rol que los partidarios del externismo estrictamente natural asignan a las propiedades subyacentes consideradas como esenciales. En el caso de rechazar esta idea (a favor de lo que argumentaremos en las siguientes secciones), habría que responder a una posible objeción de acuerdo con la cual esta variante negaría un tratamiento especial a los conceptos de género natural frente a otros términos generales que expresan conceptos susceptibles de recibir una individuación externista. Con otras palabras, el contenido de esta objeción es que para el externismo social no-esencialista *todos* los conceptos expresados por términos generales (de género natural o no) son individuados por sus condiciones de aplicación socialmente determinadas, sin que un tipo determinado de conceptos, los conceptos de género natural, tengan una condición de aplicación destacada como “esencial”, determinada por la estructura subyacente del género relevante.

4. *Por un externismo sin esencias*

En la sección anterior hemos propuesto una nueva clasificación de las distintas concepciones del externismo sobre la individuación de nuestros contenidos mentales que intenta evitar que los factores externos relevantes (naturales y sociales) entren en conflicto. Distinguimos, en este sentido, tres grandes variantes de externismo. En primer lugar, (i) un externismo estrictamente natural, que reserva a factores naturales un papel en la individuación de los conceptos expresados (exclusivamente) por términos de género natural; también (ii) un externismo de distribución, que asigna a factores

naturales y sociales un ámbito de influencia distinto para cada uno de ellos,³⁰⁴ y, por último, (iii) un externismo social. Éste, a su vez, se subdivide en un externismo social esencialista y un externismo social no-esencialista. La razón de ser de esta subdivisión descansa en las dos clases de influencia individuadora que pueden ejercer los factores del entorno natural (registrada necesariamente en los usos sociales). Así, o bien puede ligarse dicha influencia a una concepción de los géneros naturales individuados según su estructura subyacente (registrada ésta, de nuevo, en los usos sociales) o bien podemos apoyarnos en una concepción no esencialista de los mismos.

La tarea que nos aguarda hasta completar el capítulo consiste en examinar estas variedades del externismo. Comenzaremos cuestionando la primera de estas concepciones, el externismo estrictamente natural, teniendo en cuenta que su viabilidad depende fundamentalmente de la viabilidad de la teoría de los géneros naturales que este tipo de externismo pone en juego. No obstante, puesto que el externismo de distribución (en lo que respecta a los conceptos de género natural) y también, en parte, el externismo social esencialista se apoyan en la concepción de los géneros naturales mencionada, nuestras consideraciones pueden extenderse también a estas dos variedades. Nuestro fin no es, desde luego, desarrollar una concepción completa y alternativa sobre los géneros naturales y la referencia de los términos de género natural, sino más bien, destacar cuáles son los supuestos más polémicos de la concepción de los géneros naturales que respaldan los partidarios de dichas variantes de externismo y seleccionar aquellos argumentos que, a nuestro juicio, ponen en cuestión de manera más evidente sus dificultades.

Para llevar a cabo este propósito, es conveniente comenzar recordando que la intuición fundamental en la que descansan los experimentos mentales que suelen plantearse a favor (de cualquier variante) del externismo con respecto a la individuación consiste en que, al menos en ocasiones, dos hablantes situados en distintos *entornos no piensan lo mismo aunque ambos vivan en entornos idénticos por lo que a ellos respecta*.³⁰⁵ La diferencia en sus pensamientos se establece a partir de una diferencia que reside en su entorno externo, la cual determina que las palabras que profieren estos hablantes no signifiquen lo mismo. Esta diferencia semántica, a su vez, es explicada o bien por una diferencia en el uso del término que se establece en cada comunidad

³⁰⁴ Según esta variedad de externismo, los conceptos de género natural quedan individuados, en parte, por factores naturales, mientras que la identidad de los conceptos expresados por términos generales distintos de los términos de género natural quedará fijada, en parte, por factores sociales.

³⁰⁵ Esta es una expresión de Farkas (2003). Cf. Crane (2001), p. 123 y Rudd (1997), p. 506.

lingüística (individuación social), o bien por una diferencia en aquello que es designado por el término, aun cuando dicha diferencia resulte inadvertida, no ya para estos dos sujetos en particular, sino para cualquiera de los hablantes que pertenecen a su comunidad (individuación natural directa). En este último caso, ilustrado por el experimento mental de Putnam en torno al término “agua”, vemos que la lectura externista se apoya, en último término, en dos tesis adicionales, a saber, (i) que los líquidos designados por el término “agua” en la Tierra y en la Tierra Gemela son géneros naturales distintos y (ii) que esa presunta diferencia en los géneros naturales es relevante para establecer una diferencia semántica y, por tanto, una diferencia en los contenidos mentales que los protagonistas de este experimento expresan haciendo uso del término “agua”.³⁰⁶

A fin de examinar la viabilidad de estas dos últimas tesis es preceptivo estructurar el análisis de la siguiente manera. En este sentido, nos preguntaremos (a) si el escenario que plantea ese experimento es un escenario relevante para extraer conclusiones acerca la semántica de los términos de género natural, (b) si son convincentes los argumentos de Putnam según los cuales la diferente composición química de los líquidos denominados “agua” en la Tierra y en la Tierra Gemela conlleva que ambos líquidos sean géneros naturales distintos, (c) si los criterios que Putnam emplea para individuar los géneros naturales son adecuados y, por último, (d) qué aporta una respuesta positiva a estas tres cuestiones a una teoría de la individuación de los contenidos mentales que establezca una diferencia en los contenidos mentales de Óscar₁ y Óscar₂.

El examen de estos cuatro problemas nos permitirá alcanzar el doble objetivo de cuestionar aquellas concepciones del externismo que total o parcialmente se apoyan en la teoría de los géneros naturales y de la referencia de los términos de género natural de Putnam, así como examinar el potencial de aquella variedad de externismo que encuentra respaldo en otras concepciones de los géneros naturales.

4.1 *La relevancia del escenario de la Tierra Gemela*

En cuanto a la primera pregunta, puede resultar un tanto sorprendente poner en duda la relevancia del propio planteamiento de Putnam tratándose de un ejemplo filosófico tan

³⁰⁶ Sobre el papel de las intuiciones en los experimentos de Putnam remito a las críticas de Beebe/Sabbarton-Levy (2010a), p. 161 y Lycan (2006), p. 186.

reconocido y habiendo ocupado, por lo demás, el centro de la práctica totalidad de los debates sobre la presunta incompatibilidad del externismo con el acceso privilegiado.³⁰⁷ A pesar de ello, ha de reconocerse que no existe consenso sobre esta cuestión inicial, especialmente si consideramos que la relevancia de dicho experimento mental ha de estar condicionada a la posibilidad nomológica del escenario planteado por ese experimento. Nótese que en caso de sostener que este escenario no es nomológicamente posible, las pretensiones de Putnam y de todos aquellos autores que han empleado el experimento mental para extraer conclusiones sobre la semántica de los términos de género natural y defender, con ello, una concepción muy definida de los géneros naturales, serían unas pretensiones frustradas de raíz. El problema es que existen al menos dos buenas razones para pensar que ocurre de este modo y que la descripción de un mundo, la Tierra Gemela, *exactamente idéntico a la Tierra* salvo en la composición química del líquido inodoro, incoloro e insípido que llena los ríos y lagos y sacia la sed de sus habitantes, constituye un “disparate vacío” (Van Brakel 2000, p. 116).

El primer argumento que cuenta a favor de esta posición es que resulta francamente dudoso que un planeta como la Tierra Gemela, un entorno que carece de agua, sea un planeta ni siquiera remotamente parecido a la Tierra, debido a que esta sustancia no sólo es imprescindible para la supervivencia de los seres vivos de este planeta, sino que, además, figura como el componente más importante de dichos seres (en diversos porcentajes que oscilan entre el 60 y el 90%).³⁰⁸ Siendo esto así, es difícil concebir siquiera qué aspecto tendría un planeta carente de agua y cómo afectaría al desarrollo de las formas de vida allí existentes (si es que acaso resulta posible la vida sin agua). En segundo lugar, la idea de que dos sustancias pueden tener propiedades microestructurales tan distintas como Putnam pretende y que, sin embargo, posean las mismas propiedades manifiestas – así como unas propiedades físicas idénticas, como su densidad, punto de ebullición, etc. – atenta contra uno de los rasgos que este autor atribuye a la estructura subyacente de los géneros, a saber, que ha de ser causalmente responsable de las propiedades manifiestas o macroestructurales. En este sentido, el líquido relevante de la Tierra Gemela es, si no una imposibilidad, al menos una

³⁰⁷ En los artículos dedicados a discutir la posibilidad de compatibilizar el externismo y el acceso privilegiado son casi anecdóticas las páginas que se han dedicado a los demás experimentos mentales o al resto de variantes del externismo. En nuestra opinión, esto no es una casualidad puesto que, según veremos, el experimento mental en torno al término “agua”, así como el externismo estrictamente natural que sustenta es especialmente agresivo con respecto al acceso privilegiado a los propios contenidos mentales.

³⁰⁸ Cf. Monge-Nágera/Gómez/Rivas (2002), p. 54.

aberración nomológica de primer orden que contraviene la propia concepción de los géneros naturales y sus propiedades ideada por Putnam.³⁰⁹

Los dos argumentos ofrecidos sirven de justificación a aquellos autores que han desechado cualquier relevancia que el escenario de Putnam puede tener para nuestra concepción de los géneros naturales y la semántica de los términos de género natural, mostrando con ello una actitud escéptica sobre lo que cabría decir en una situación fantástica como la recreada por el experimento mental de la Tierra Gemela (Cf. Dupré 1981, p. 71).³¹⁰

4.2 Cuestionamiento del caso particular del agua

Existe una variedad de respuestas ante la cuestión de si cabe aceptar, con Putnam, que la diferente composición química que se estipula en los líquidos denominados “agua” en la Tierra y en la Tierra Gemela hace que éstos deban ser considerados géneros naturales distintos. Como vimos, los apoyos teóricos que Putnam pone en juego para dar una respuesta afirmativa son, en primer lugar, la identificación de la estructura subyacente del agua con su composición química, H_2O , y, en segundo lugar, una teoría de los géneros naturales que hace que la estructura subyacente que presuntamente comparten todos los ejemplares de un género natural sea el elemento que determina la extensión de éste. Así, en la medida en que aceptemos que la propiedad de tener la composición química H_2O es determinante para considerar que una porción de líquido es agua, se seguirá de ello que las porciones del líquido de la Tierra Gemela *no son agua*, al tener una composición química distinta. Hecha esta precisión, en esta subsección nos ocuparemos de discutir solamente el caso particular del agua, que es el género natural en torno al cual se plantea el único de los experimentos mentales externistas clásicos que presta apoyo a la variante estrictamente natural del externismo. En particular,

³⁰⁹ Esta observación puede encontrarse en Searle (1983), p. 203, Segal (2000), p. 25, Wikforss (2010), p. 77 y Dupré (1981), p. 72. También en Hallett (1991, p. 20) leemos: “Como [el propio Putnam] advierte, la estructura oculta parece importante debido a su conexión con las características superficiales (gravedad específica, fluidez, punto de congelación, sabor, sensación táctil característica, etc.). Pero su ficción de la Tierra Gemela corta esa conexión; la diferencia en la microestructura no determina ninguna diferencia perceptible en ambos líquidos [...] Por alguna extraña razón, la disimilitud estructura entre el H_2O y el XYZ podría tener incluso menos consecuencias que entre el H_2O (agua ordinaria) y el D_2O (agua pesada). El término ‘agua’ podría por tanto ser extendido al líquido de la Tierra Gemela de Putnam como si en realidad fuera, por ejemplo, agua pesada”.

³¹⁰ Hanfling se pronuncia en el mismo sentido: “en filosofía es útil, con frecuencia, introducir experimentos mentales que están más allá de los límites de la posibilidad empírica; pero estos no deberían ir más allá de los límites del sentido” (Hanfling 2000, p. 233).

examinaremos aquellas razones que se han aducido para cuestionar que la composición química H_2O constituye por sí misma el criterio para determinar qué es agua y qué no lo es.

En primer lugar, hay una serie de datos que nos instan a rechazar la idea de que la posesión de dicha composición química constituye una *condición suficiente* para que una sustancia sea agua. En este sentido, cabe decir que los líquidos contenidos, respectivamente, en una taza de té y en una lata de *Sprite* contienen una proporción de H_2O prácticamente idéntica a la que contiene el agua que sale de nuestros grifos, aun cuando nadie esté dispuesto con seriedad a llamar “agua” al té o al *Sprite*.³¹¹ En segundo lugar, en cuanto a la propuesta de que la propiedad de estar compuesto de H_2O constituye una *condición necesaria* para que un líquido sea agua, también existen serios problemas al respecto, especialmente si tenemos en cuenta que la especie molecular H_2O constituye sólo uno de los dieciocho compuestos formados a partir de tres isótopos del hidrógeno (h-1, h-2, h-3) y del oxígeno (o-16, o-17, o-18) y que, además, las porciones de H_2O (compuesto de hidrógeno-1 y oxígeno-16) no son las únicas porciones compuestas de hidrógeno y oxígeno que reciben el nombre de “agua”. De este modo, también el HDO,³¹² el D_2O ,³¹³ el HTO³¹⁴ y el T_2O ,³¹⁵ otras porciones que constan de moléculas compuestas de los mismos elementos que el H_2O , se denominan, respectivamente, “agua semipesada”, “agua pesada”, “agua semi-superpesada” y “agua superpesada”. Esta información serviría para persuadirnos de que, desde un punto de vista científico, no es totalmente desatinado considerar que el XYZ es *otra clase de agua*, debido a que la ciencia reconoce, de hecho, distintas clases de agua. Si aceptáramos esta consecuencia, la conclusión de que los habitantes de la Tierra y de la Tierra Gemela se refieren a géneros naturales distintos mediante el uso del término “agua” se vería debilitada.³¹⁶

Cabe poner en duda, además, que el líquido que habitualmente denominamos “agua” (a secas) sea (sólo) H_2O . El argumento fundamental a este respecto es que el agua que bebemos, que sale del grifo, que cae del cielo en forma de lluvia, que llena ríos

³¹¹ Cf. Chomsky (1995), p. 42. Igualmente, Abbott en (1999, p. 145) destaca que los pollos, los bebés y los tomates contienen una mayor proporción de H_2O que el Gran Lago Salado de Utah.

³¹² Compuesto de un átomo de oxígeno-16, un átomo de hidrógeno-1 y un átomo de hidrógeno-2 (o deuterio).

³¹³ Compuesto de un átomo de oxígeno-16 y dos deuterios.

³¹⁴ Compuesto de un átomo de oxígeno-16, un átomo de hidrógeno-1 y un átomo de hidrógeno-3 (o tritio).

³¹⁵ Compuesto de un átomo de oxígeno-16 y dos tritios.

³¹⁶ Así, se pregunta Castañeda, “¿[p]or qué no podría el XYZ, que compone supuestamente el agua gemela pura, ser una variedad molecular más del agua?” (Castañeda 1989, p. 116).

y lagos y que *es, incuestionablemente, agua*, contiene trazas de otras moléculas formadas a partir del oxígeno y del hidrógeno, distintas del H₂O. La explicación de este hecho reside en que, aunque el hidrógeno que suele encontrarse en la naturaleza es, fundamentalmente, hidrógeno-1, también contiene pequeñas porciones de deuterio y tritio, todos ellos con un comportamiento químico y termodinámico similar, por lo que a menudo sucede que los procesos químicos naturales que dan lugar a la formación de H₂O (predominantemente) también conllevan la formación de otros isotopólogos del hidrógeno y el oxígeno (Weisberg 2006, p. 240). Por este motivo, es relativamente frecuente que muchas porciones del agua de nuestro entorno no estén compuestas sólo de moléculas de H₂O, sino que también contengan trazas de D₂O, lo cual nos indica que la composición del género natural agua es más heterogénea de lo que Putnam supone en su experimento.

Las observaciones precedentes nos invitan a pensar que si deseáramos mantener que una sustancia es agua si y sólo si su composición química es H₂O, tendríamos que admitir las siguientes consecuencias, a saber, o bien dejamos de considerar que el líquido que llena nuestros embalses y sale de nuestros grifos es agua, o bien obviamos el hecho de que hay otros isotopólogos del hidrógeno y del oxígeno distintos del H₂O que también reciben el nombre de “agua”,³¹⁷ o bien admitimos que el agua tiene una composición heterogénea. Ahora bien, la primera opción resulta extremadamente contraintuitiva, puesto que habíamos convenido en que ese líquido era, indudablemente, agua. La segunda supone pasar por alto usos científicos bien establecidos del término “agua” y, en lo que respecta a la última alternativa, hemos visto que su adopción resultaría fatal para la lectura externista del escenario de la Tierra Gemela, ya que, de aceptar que el agua no es sólo H₂O, sino que tiene una composición química heterogénea, tendríamos motivos para considerar que el XYZ es (un tipo de) agua y, con ello, resistirnos a la conclusión de que los líquidos de un entorno y otro son géneros naturales distintos.

Llegado este punto, podemos extraer la siguiente conclusión. *Ni siquiera hoy*, cuando nuestro bagaje de conocimiento científico nos permite conocer la diferencia existente entre el agua y el líquido relevante de la Tierra Gemela, es una postura incontrovertible afirmar que el agua es (sólo) H₂O y que el líquido de la Tierra Gemela

³¹⁷ No faltan autores que han señalado que a partir de la concepción de los géneros naturales de Putnam sería más razonable considerar que el D₂O y el H₂O son géneros naturales distintos (Van Brakel 2000, p. 107).

no es agua de ninguna manera. Si apelando al desarrollo científico actual no está claro que podamos establecer una diferencia verdaderamente determinante entre estas sustancias, una diferencia que nos convenza de que el XYZ no es (ninguna clase de) agua, cuánto más cuestionable será proyectar esa diferencia hasta el tiempo en que viven Óscar₁ y Óscar₂, en el cual no existía tal desarrollo, ni la consiguiente posibilidad de detectar diferencia alguna entre ambos líquidos y, sobre todo, cuando esa supuesta diferencia entre los conceptos que se expresan en cada entorno mediante el término “agua” no tiene el más mínimo efecto en la comunicación o en la competencia lingüística (Lycan 2006, p. 186).

4.3 *La viabilidad de la concepción de los géneros naturales del externismo estrictamente natural*

No obstante, a pesar de estas objeciones concernientes al caso particular del género agua, un partidario del externismo estrictamente natural puede replicar de la siguiente manera. Si bien el género natural escogido en el experimento de Putnam parece especialmente problemático, podría plantearse un nuevo experimento en torno al uso de otro término de género natural que no presente esta clase de dificultades. En cualquier caso, se nos diría, lo que merece la pena ser discutido no es tanto el ejemplo particular en el que encuentra apoyo el experimento mental de la Tierra Gemela, sino, más bien, los supuestos de la formulación de todos esos ejemplos. De aceptar el reto que plantea esta objeción, tendríamos que hacernos cargo de aquellos supuestos operativos en la lectura externista del escenario de la Tierra Gemela y que podrían servir para construir otros escenarios semejantes en los que, a partir de una diferencia (completamente desconocida para *todos* los hablantes) en las propiedades subyacentes de cualesquiera dos géneros naturales de dos entornos distintos, tendríamos que establecer una diferencia correspondiente en el contenido de los pensamientos de aquellos sujetos que, en cada entorno, pretenden referirse a los ejemplares de cada uno de estos géneros.

Siguiendo esta sugerencia, vamos a enumerar y analizar estos supuestos:

- (1) Un género natural viene determinado por la posesión por parte de sus miembros de una misma naturaleza o estructura subyacente (Putnam 1975c, pp. 232-233).
- (2) En los usos ordinarios y pre-científicos de los términos de género natural se presupone la existencia de esa naturaleza compartida entre los ejemplares del

género. Los hablantes tienen ciertas “intenciones referenciales” al usar los términos de género natural. En particular, su intención es que una entidad pertenece a la extensión de un término de género natural si y sólo si posee las mismas propiedades subyacentes que los miembros paradigmáticos del género. Tales intenciones tienen su fundamento, según Putnam, en el hecho de que sabemos que hay conjuntos de entidades con una “estructura oculta común” aun cuando no sepamos todavía cómo describir tales estructuras (1975c, p. 244).

- (3) La estructura subyacente compartida por los ejemplares de un género natural es descubierta (no estipulada) por la actividad científica (Putnam 1975c, p. 223).³¹⁸
- (4) Debido a la existencia en todos los hablantes de la clase de intenciones referenciales descritas en (2), los hablantes profanos están dispuestos a rectificar su uso de los términos de género natural a partir del dictamen de los expertos, en el sentido de que si se descubre que una serie de entidades que, inicialmente, se consideraban pertenecientes a un mismo género natural forman parte, en realidad, de dos géneros naturales distintos *debido a que poseen propiedades subyacentes distintas*, los hablantes profanos rectificarán su uso de los términos de género natural relevantes a partir de dicho descubrimiento (1975c, p. 223).³¹⁹

Estos son los supuestos que sustentan la tesis de que el significado de un término de género natural “T” queda determinado, en parte pero fundamentalmente, por la naturaleza subyacente del género – aun cuando ésta no haya sido descubierta – y que permitirían construir experimentos similares al experimento de la Tierra Gemela empleando otros géneros naturales distintos del agua. De este modo, a partir de una diferencia en las propiedades subyacentes de las entidades denominadas “T” en dos entornos distintos, deberíamos establecer una diferencia semántica en “T” cuando es usado por los hablantes bien de un entorno, bien de otro, si aceptamos que los habitantes de ambos entornos presuponen la existencia de una naturaleza subyacente compartida por todas las instancias que denominan “T” y si aceptamos, además, que esa naturaleza

³¹⁸ Vid. Putnam (1970), p. 140.

³¹⁹ Esta tesis constituye uno de los supuestos de la primera parte del experimento mental de la Tierra Gemela en torno al uso del término “agua” descrita en la nota 65 del presente trabajo. De este modo, los viajeros terrestres que llegan a la Tierra Gemela suponen inicialmente que el XYZ es agua, dado el parecido superficial de ambas sustancias, pero “cuando se descubriera que ‘agua’ en la Tierra Gemela es XYZ, esta suposición se rectificaría” (Putnam 1975c, p. 223).

es susceptible de ser descubierta por sus comunidades científicas como aquella que *desde siempre* ha constituido al género.

Una vez desglosados los supuestos operativos en las tesis del externismo estrictamente natural, vamos a ocuparnos de cada uno de ellos y a ofrecer, correspondientemente, una serie de argumentos para defender que cualquiera de ellos dista de ser un supuesto evidente acerca de los géneros naturales y de la semántica de los términos de género natural.

Comenzaremos, pues, con el supuesto (1), a saber, que un género natural viene determinado por la posesión por parte de sus miembros de una misma naturaleza o estructura subyacente. Lo primero que ha de señalarse es que este supuesto recoge propiamente la clase de esencialismo sobre los géneros naturales que subyace al externismo estrictamente natural. A nuestro juicio, el problema más grave que ha de afrontar quienquiera que respalde este supuesto sobre los géneros naturales es que se trata de una concepción de los mismos muy difícil de encajar en el ámbito de la biología. Si estuviéramos en lo correcto y existieran dificultades de peso para considerar que las especies animales y vegetales conforman géneros naturales a partir de una naturaleza intrínseca compartida, entonces este problema resultaría ser especialmente grave para los partidarios de este supuesto, ya que los términos de especies biológicas y botánicas son considerados, de manera paradigmática, como términos de género natural.³²⁰

El problema básico para aceptar que esta visión esencialista de los géneros naturales resulta apropiada a la hora de dar cuenta de ese tipo de géneros naturales que son las especies es que actualmente no hay un consenso entre los propios biólogos en torno a la cuestión de cuál es el concepto de especie válido para clasificar a los organismos vivos,³²¹ sino que, más bien, coexisten varios conceptos que, a partir de un criterio de individuación destacado, dan lugar a distintas categorizaciones. La concepción de lo que conforma una especie y la clasificación biológica asociada variarán, por tanto, según se considere que el criterio fundamental es la descendencia de un ancestro común, el aislamiento reproductivo entre individuos de una misma

³²⁰ En este sentido, es digno de destacarse que una gran cantidad de autores se han mostrado especialmente críticos con la aplicación de la teoría de los géneros naturales de Putnam y Kripke al ámbito de los seres vivos. Así, Dupré resume su crítica de una manera muy expresiva diciendo que “la historia de Putnam no funciona en biología” (Dupré 2002, p. 3) e igualmente Okasha (2002, p. 198), quien afirma que “las teorías de Putnam y Kripke entran en conflicto con las opiniones recibidas en biología”. Véase también Bhushan (2006), p. 331 y Sober (2000), pp. 148 y ss.

³²¹ El artículo de Mayden de (1997), en el que se recogen hasta 22 conceptos de especie operativos en la biología actual, resulta especialmente revelador al respecto.

especie,³²² la adaptación a un entorno natural específico,³²³ las relaciones filogenéticas,³²⁴ etc.

En segundo lugar, no sólo ocurre que el uso de un criterio u otro produce distintas categorizaciones de las entidades biológicas, sino que hay razones para pensar que algunos de ellos sólo son efectivos con respecto a un grupo restringido de organismos. De este modo, los organismos procariotas (carentes de un núcleo bien definido) resultan especialmente difíciles de clasificar en especies diferenciadas, cualquiera que sea el criterio utilizado para ello.³²⁵ La relevancia de esta observación puede percibirse cuando reparamos en que estos organismos son los más abundantes de la tierra, lo cual supone, en último término, que existen serias dificultades para elaborar un único concepto de especie operativo para la mayoría de los seres vivos.³²⁶

Ante las dificultades que encuentra en el ámbito de la biología la teoría de los géneros naturales que respalda el externismo estrictamente natural, una de las concepciones de los géneros naturales que cuenta con más apoyo filosófico en la actualidad dista mucho de ser tan restringida como la que pone en juego la teoría mencionada. En este sentido, en los últimos años parece haberse aceptado que, en caso de que una teoría de los géneros naturales quiera tener alguna oportunidad en el ámbito de las especies biológicas, resulta imprescindible que las propiedades individuadoras de las especies sean entendidas más como propiedades funcionales realizadas en organismos intrínsecamente heterogéneos que como una naturaleza interna subyacente

³²² En este sentido, cabe indicar con respecto a muchos seres vivos que tienden a reproducirse con individuos de la misma especie y, además, los organismos híbridos, cuyos progenitores pertenecen a especies distintas, suelen ser estériles.

³²³ Este criterio se basa en el hecho de que cada nicho ecológico determina que los organismos que comparten una historia evolutiva desarrollen unos mecanismos de adaptación similares.

³²⁴ En una clasificación en base a criterios filogenéticos se seleccionan una serie de datos genéticos (el ADN ribosómico, la hibridación del ADN, etc.) como los más relevantes para la constitución de la especie (Ereshefsky 2010a, p. 556).

³²⁵ Ereshefsky en (2010a) sostiene que el problema que existe con esta clase de entidades biológicas es que su proceso reproductivo varía considerablemente con respecto al que caracteriza a los seres vivos más complejos, ya que por lo general intercambian su material genético sólo parcialmente y, además, las partes intercambiadas en el proceso pueden dar lugar con rapidez a historias evolutivas muy diferentes. Un proceso reproductivo de esta clase implica que (i) su aislamiento reproductivo es mínimo, especialmente si lo comparamos con el de otros organismos más complejos; (ii) la rapidez y amplitud de estas combinaciones genéticas parciales hace de la identidad del nicho ecológico un criterio de clasificación demasiado amplio y (iii) tampoco existe una estabilidad significativa en los datos genéticos que, desde un enfoque filogenético, se emplearían como criterio de determinación de la especie. En el mismo sentido cf. Ereshefsky (2010b), p. 680.

³²⁶ Para dar una idea de la importancia de los organismos procariotas hay que tener en cuenta que integran dos reinos animales (de seis reinos existentes), los *Archaeobacteria* y *Eubacteria*. Las bacterias que pertenecen al primer grupo constituyen las formas de vida más antiguas conocidas y son los organismos más simples desde un punto de vista biológico. Aunque tanto las arqueobacterias como las eubacterias carecen de núcleo, las primeras son a nivel molecular más semejantes a los organismos eucariotas que las segundas.

que caracteriza unívocamente a todos (y sólo a) los miembros de un género. Constituye un buen ejemplo al respecto la teoría del Racimo Homeostático de Propiedades o teoría (RHP), desarrollada principalmente por Boyd,³²⁷ que nos propone individuar los géneros naturales formados por organismos vivos a partir de aquel conjunto de propiedades que, por lo general, concurren en un conjunto de entidades. La coexistencia de esas propiedades en los seres que pertenecen a un género natural se debe a la existencia de mecanismos subyacentes causalmente responsables de una suerte de “equilibrio homeostático” en el que permanecen aquellas propiedades incluidas en el racimo. El concepto de homeostasis se emplea en biología para referirse al conjunto de procesos que permiten mantener el equilibrio de las actividades internas y funciones de un ser vivo. Los partidarios de la teoría (RHP) hacen un uso metafórico de este concepto para dar a entender, con ello, que los mecanismos subyacentes son responsables del equilibrio que existe entre las propiedades que componen el racimo, dando como resultado que la mayor parte de los miembros de un género natural poseen un subconjunto de esas propiedades (aunque no todas ellas necesariamente) y que las propiedades idiosincrásicas de alguno/s de esos miembros tienden a desaparecer a favor de las propiedades que componen el racimo.

A pesar de que esta teoría sobre los géneros naturales presenta por sí misma un gran interés, lo más relevante para nuestro propósito es extraer las siguientes conclusiones. En primer lugar, las concepciones de los géneros naturales desarrolladas en los últimos años, como la propia teoría (RHP), otorgan poca plausibilidad a la exigencia de una naturaleza subyacente intrínseca, común a los miembros de un género, en tanto que condición necesaria y suficiente de pertenencia al mismo.³²⁸ En segundo lugar, desde una teoría de los géneros naturales como (RHP) el planteamiento de escenarios similares al diseñado en el experimento mental de la Tierra Gemela se vuelve mucho más problemático. El motivo es que a partir de la visión esencialista de los géneros naturales que respalda esta clase de externismo – recogida en (1) – es más verosímil que una variación contrafáctica en la propiedad esencial de un género natural conlleve una variación correlativa en el significado del término natural relevante y, en último término, una variación en el contenido de aquellos pensamientos en cuya

³²⁷ Cf. Boyd (1999a) y (1999b), así como Wilson/Barker/Brigandt (2007).

³²⁸ De hecho, los seguidores de la teoría (RHP) mantienen que una de las grandes virtudes de sus propuestas es que permiten que organismos que pertenecen a un género natural sean intrínsecamente heterogéneos, algo impensable desde el punto de vista de la teoría de los géneros naturales que subyace al externismo estrictamente natural (Cf. Wilson 1999b, pp. 190 y ss.).

expresión figura ese término. Sin embargo, si aceptamos, como pide la teoría (RHP), que la constitución de un género natural viene determinada por un racimo de propiedades, resulta más complicado determinar cuántas de esas propiedades han de variar contrafacticamente para considerar que se ha producido un cambio en el significado del término de género natural relevante y, consiguientemente, el procedimiento para extraer unas conclusiones favorables al externismo estrictamente natural se vuelve más indirecto y complejo.

No obstante esto, un defensor de esta clase de externismo podría argüir que, aunque su tesis presenta ciertas fricciones a la hora de ser acomodada al ámbito de los seres vivos, sigue siendo perfectamente adecuada para explicar la individuación de elementos químicos, como el oro o el aluminio, o de compuestos químicos como el agua. Ante esta alegación se hace preciso reconocer que la teoría de los géneros naturales que respalda el externismo estrictamente natural se acomoda relativamente bien al caso particular de los elementos químicos. La razón es que el número atómico, esto es, el número de protones en el núcleo de un átomo, cumple al parecer el rol que Putnam reserva a las propiedades subyacentes de un género, en cuanto que (i) es causalmente responsable de la estructura atómica del elemento y de su comportamiento químico y (ii) sirve como criterio de individuación de cada uno de los elementos.³²⁹

Sin embargo, en el caso de los compuestos químicos la adaptación no es tan sencilla. En primer lugar, con relación a esa sustancia que es el agua ya hemos comprobado que determinar la presunta estructura subyacente (de carácter químico) que constituye su extensión no es tarea fácil. En este sentido, no se trata sólo de que la descripción que Putnam ofrece de la esencia de este género natural resulte extremadamente simplista, sino que existen dificultades de principio para determinar una única propiedad como *la* propiedad esencial de dicho género. Así, por un lado, se ha señalado que las propiedades relevantes para individuar la extensión del agua no son únicamente las propiedades químicas (como su composición molecular),³³⁰ sino que las *propiedades funcionales* también podrían servirnos a este respecto y lo mismo podría ocurrir con otra clase de compuestos químicos. En el caso particular del agua, dado el

³²⁹ Decimos que la acomodación de esta teoría al caso de los elementos químicos es sólo *relativamente* buena porque, aun cuando el número atómico de los elementos cumple estos dos rasgos que Putnam atribuye a la estructura subyacente, cabe cuestionar que el proceso de fijación de la referencia de muchos elementos químicos haya seguido las pautas que Putnam describe. Como veremos en las próximas páginas, Beebe y Sabbarton-Leary analizan en (2010a) el caso de ciertos elementos químicos cuya referencia se ha fijado estipulativamente.

³³⁰ Cf. LaPorte (2010), p. 110.

carácter imprescindible de esta sustancia para la supervivencia y composición de todas las formas conocidas de vida, tanto la potabilidad como aquellas funciones biológicas en que se ve involucrada constituirían propiedades funcionales especialmente relevantes para caracterizar esta sustancia y determinar, con ello, qué porciones de líquido pertenecen al género. De aceptar esta sugerencia, tendríamos una razón adicional para considerar que el agua y el líquido relevante de la Tierra Gemela son dos tipos de agua, más bien que dos géneros naturales distintos, visto que ambas sustancias poseen unas propiedades funcionales similares.

Por otro lado, no sólo esta clase de propiedades podría servir para determinar la extensión de un compuesto químico sino que algunas *propiedades físicas* (como el punto de ebullición, la densidad mínima o la viscosidad de una sustancia) podrían emplearse como criterios de individuación de la sustancia en cuestión. De hecho, no han faltado autores que sostienen que cada una de estas propiedades puede servir tan bien como la composición molecular para individuar la extensión de una sustancia.³³¹ Algunas de las razones que aportan para defender esta idea son las siguientes. En primer lugar, todas ellas se encuentran interrelacionadas con la composición molecular, lo cual constituye un excelente motivo para considerar que se trata de propiedades al menos tan centrales como esta última. En segundo lugar, se trata de propiedades cuyo conocimiento está, en ocasiones, relativamente extendido entre los hablantes y que, en tercer lugar, pueden ser adecuadas para vincular los estereotipos asociados al término con los tests empleados por los expertos. Estas dos últimas características permiten, además, reducir los problemas que, según veremos, suscita la distinción entre los usos ordinarios de un término de género natural y los usos científicos del mismo. En cuarto y último lugar, aunque existiera un cambio en las teorías científicas de tal forma que descubriéramos que la composición molecular de una sustancia – considerada como la estructura subyacente – era errónea, es muy probable que dicho cambio no afectase al valor del punto de ebullición de la misma, dado lo cual dicha propiedad podría emplearse para explicar la continuidad referencial del término a través de las distintas teorías.

Las consideraciones precedentes sobre la relevancia de las propiedades funcionales y físicas de las sustancias químicas aportan plausibilidad a la idea de que la composición molecular no debería ser considerada como la propiedad esencial de una

³³¹ Cf. Van Brakel (2000), pp. 116 y ss., Needham (2000), p. 21 y Weisberg (2006), p. 342.

sustancia química, sino que, por el contrario, existen una serie de propiedades especialmente relevantes (e interrelacionadas entre sí) que pueden ser empleadas para determinar el género.

Pasemos a ocuparnos ahora del supuesto (2) de la teoría de los géneros naturales que suscriben los partidarios del externismo estrictamente natural, supuesto que, como indicamos, afirma que en los usos ordinarios y pre-científicos de un término de género natural presuponemos la existencia de una naturaleza compartida por los ejemplares del género.³³² Para abordar su análisis conviene hacer una observación preeliminar y distinguir en el seno de los usos ordinarios de los términos de género natural aquellos usos que preceden a los supuestos descubrimientos de esencias (usos pre-científicos del término) y, por otro, los usos de los términos de género natural por parte de profanos una vez que dichos descubrimientos han tenido lugar.

El desarrollo que los partidarios de la concepción de los géneros naturales que respalda esta clase de externismo efectúan del supuesto (2) puede articularse de la siguiente manera. Para estos autores las intenciones referenciales mencionadas sustentan la existencia de una continuidad, en primer lugar, entre los usos pre-científicos de un término de género natural y los usos del término que tienen lugar después del descubrimiento de la estructura subyacente del género, ya que los hablantes que llevan a cabo estos usos pre-científicos pretenden usar el término para designar entidades que comparten las mismas propiedades subyacentes, el descubrimiento de las cuales podrá producirse en el futuro. En segundo lugar, estas intenciones referenciales conllevan, en relación a los hablantes profanos que llevan a cabo usos del término posteriores al descubrimiento de las propiedades subyacentes, una actitud de deferencia hacia el dictamen de los especialistas presentes en su comunidad científica actual, lo que determina, en último término, que exista también una continuidad entre el concepto que se expresa en los usos ordinarios y el concepto expresado en los usos científicos, en virtud del hecho de que los hablantes profanos confían en el criterio de los expertos de su comunidad a la hora de determinar la referencia de los términos de género natural a partir de su conocimiento de las propiedades subyacentes.³³³

³³² Como puede apreciarse, el supuesto (2), que reza que en los usos ordinarios y pre-científicos de un término de género natural presuponemos la existencia de una naturaleza compartida por los ejemplares del género, esta relacionado con el supuesto (1), a saber, que un género natural viene determinado por la posesión por parte de sus miembros de una misma naturaleza o estructura subyacente. Debido a esto, algunas de las críticas que en lo sucesivo dirigiremos al supuesto (2) también afectarán al supuesto (1).

³³³ Tanto Weisberg en (2006) como Dupré en (2002, p. 43) han señalado que la concepción de los géneros naturales que respalda el externismo estrictamente natural asume la presencia de una continuidad basada

Sin embargo, también en este caso es posible ofrecer algunos argumentos contrarios a esta idea de que el hablante ordinario que usa un término de género natural (tanto antes como después del descubrimiento de las propiedades subyacentes de un género) tiene las intenciones referenciales mencionadas. Así, en primer lugar, cabe pensar que los hablantes profanos, incluyendo aquellos que viven en el tiempo posterior al presunto descubrimiento de las propiedades subyacentes, con frecuencia usan un término de género natural sin pretender referirse a ejemplares de sustancias químicas que, desde el punto de vista de su composición, conforman un único género natural; en cambio, el uso de esta clase de términos por parte de los expertos relevantes es extremadamente sensible al contexto y mucho más fino y específico que el uso que llevan a cabo los hablantes ordinarios.³³⁴

En segundo lugar, tanto la presencia de tales intenciones referenciales como la actitud de deferencia que se fundamenta en ellas pueden cuestionarse de una manera distinta partiendo de la asunción de que, si dichas intenciones existieran de manera generalizada, las clasificaciones ordinarias de los géneros naturales tenderían a coincidir con las clasificaciones científicas de los géneros, suponiendo que la ciencia es la actividad encargada de revelar aquella propiedad que, según se alega, determina la constitución de los géneros. Sin embargo, podemos apreciar que tal coincidencia no tiene lugar y que existe una gran disparidad entre las clasificaciones ordinarias y las científicas por varios motivos, entre los que cabe señalar los siguientes. Por una parte, las clasificaciones ordinarias de géneros naturales están, por lo general, restringidas a seres vivos macroscópicos y sustancias presentes en el entorno natural más inmediato, mientras que las clasificaciones científicas no están afectadas por esta clase de limitaciones. Por otra parte, la denominación y distinción de géneros naturales en contextos ordinarios se rige por intereses muy variados que, en ocasiones, atienden a

en la naturaleza común a los ejemplares de cada género. Mientras que Weisberg habla al respecto de “un principio de coordinación”, Dupré enuncia una tesis similar, aunque no emplea una denominación específica para la misma: “Putnam trató las clasificaciones ordinarias como primeras aproximaciones a taxonomías científicas y afirmó que era algo inherente a nuestro uso ordinario de términos de género natural que ellos están sujetos a refinamiento en cuanto las taxonomías científicas fueran desarrolladas y mejoradas [...] Este hecho no es en absoluto el caso generalmente”. LaPorte también ha cuestionado esta clase de continuidad al afirmar que los usos de los términos de género natural previos a aquellos descubrimientos científicos que han señalado una propiedad como esencial de un género natural son usos de “textura abierta” (2010, p. 117), es decir, aplicaciones del término sin límites completamente definidos (2004, p. 100). También en Dupré (1993, p. 24) encontramos esta observación.

³³⁴ “[E]stos términos [de género natural aplicados a las sustancias químicas] seleccionan diferentes géneros químicos en diferentes contextos conversacionales y de explicación. El término ‘agua’, usado por un químico, a veces se referirá a los isómeros de H₂O en la naturaleza, otras veces a otro isómero de H₂O y, a veces, quizá, a un ejemplar homogéneo de H₂¹⁶O, dependiendo de las circunstancias de la preferencia” (Weisberg 2006, p. 342).

rasgos de las entidades naturales que, desde el punto de vista científico, carecen de relevancia. En este sentido, el interés sociológico, el asombro intelectual que suscita una clase de animales o plantas, la familiaridad que se establece con ellos, la importancia económica de estas entidades, etc., son factores que determinan la introducción en el lenguaje de términos de género natural, así como su uso en contextos ordinarios.³³⁵ Puede encontrarse un buen ejemplo de esta asimetría entre los usos ordinarios y los usos científicos de los términos “carbón” y “diamante” (LaPorte 2004, pp. 100-101). Asimismo, los trabajos de Dupré contienen una gran cantidad de observaciones de este tenor, como el dato de que los ajos y las cebollas pertenecen a una misma especie botánica (y, por tanto, pertenecerían a un único género natural), aun cuando la diferencia entre una y otra clase de plantas es extremadamente importante desde muchos puntos de vista en contextos ordinarios.³³⁶

Desde luego, ante este argumento un partidario del externismo estrictamente natural podría ofrecer la siguiente réplica. Quizá nos concedería que muchos términos que pretenden designar géneros naturales en contextos ordinarios (pre y post-científicos) no tienen un equivalente en contextos científicos y, al revés, que los intereses científicos impulsan la creación e introducción en el lenguaje de una serie de términos de género natural que no son usados por los hablantes ordinarios. Sin embargo, puesto que la ciencia tiene como tarea revelarnos las propiedades subyacentes de los seres naturales y, por tanto, se espera de ella que sea capaz de establecer la taxonomía que nos permita “cortar la naturaleza en sus articulaciones”,³³⁷ las clasificaciones ordinarias tienen una validez subordinada a los dictámenes de la ciencia. Por este motivo, si un término que en los usos ordinarios pretende aplicarse a un género natural no designa, finalmente, entidades dotadas de una misma naturaleza, entonces debemos concluir que no se trata, en realidad, de un término de género natural genuino.³³⁸ Esto ocurriría en dos tipos de casos. En primer lugar, si la ciencia descubriera que un presunto término de género natural no designa un conjunto de entidades con una misma naturaleza subyacente,

³³⁵ Cf. Dupré (1993), p.34 y (2002), p. 31.

³³⁶ Lo mismo ocurre con las ranas y los sapos, los (pájaros) herrerillos y los carboneros, el repollo y la coliflor (Dupré 2002, p. 45).

³³⁷ Esta frase alude a un pasaje del diálogo platónico *Fedro* (265d-266a) en el que se establece la meta del proceder filosófico del siguiente modo: “dividir las ideas en clases, siguiendo sus articulaciones naturales, sin romperlas por cualquier parte, tal y como lo haría un mal carnicero”. El objetivo de las teorías realistas sobre los géneros naturales con frecuencia se describe cambiando el pasaje platónico por la oración citada en el cuerpo del texto, “cortar la naturaleza en sus articulaciones”, es decir, descubrir cuál es la clasificación que, de suyo, presentan los seres naturales.

³³⁸ Bird mantiene esta posición en (2010).

dicho término debería dejar de ser considerado un término de género natural (el caso, curiosamente, del término “haya”, que designa distintas especies botánicas). En segundo lugar, también sucedería lo mismo si ciertas diferencias entre los seres naturales que se juzgan relevantes a la hora de introducir dos términos de género natural en el discurso ordinario no reflejaran, en realidad, una diferencia de naturaleza subyacente (el caso de “carbón” y “diamante”, “topacio” y “rubí” o “cebolla” y “ajo”).

No obstante, estimamos que esta solución es sumamente radical y para justificar nuestra opinión, poco partidaria de actitudes revisionistas de esta clase, ofreceremos dos argumentos. El primero de ellos se basa en la creencia de que existe una tensión entre los supuestos de esta teoría de los géneros naturales y la práctica efectiva de la ciencia actual. Nuestro razonamiento puede presentarse en forma de trilema, mostrando, con ello, que los partidarios de dicha teoría tendrían que asumir tres tesis mutuamente inconsistentes. En primer lugar, tales autores parecen aceptar, en principio, que (a) los términos que designan especies biológicas son, paradigmáticamente, términos de género natural según la concepción más usual de esta clase de expresiones; en segundo lugar, su teoría de los géneros naturales defiende que (b) la ciencia es la (única) actividad encargada de determinar las propiedades que constituyen los géneros naturales, en detrimento de otras actividades regidas por intereses distintos, y, por último, deberían explicar el dato de que (c) en el seno de la propia ciencia biológica se acepta la existencia de una variedad de caracteres genéticos, morfológicos o fisiológicos dentro de una misma especie de seres vivos.³³⁹ Esta última tesis, como vimos, supone cuestionar que aquellos seres vivos que la biología considera como seres pertenecientes a una misma especie tengan todos ellos una naturaleza intrínseca homogénea, tal y como pretende esta concepción de los géneros naturales.

Teniendo en cuenta estas ideas, vemos que cuando un defensor del supuesto (2) persiste en mantener esta teoría de los géneros naturales y sólo acepta, en tanto que géneros naturales genuinos, aquellas especies biológicas cuyos miembros poseen una misma estructura subyacente, ha de hacer frente al hecho establecido de que la propia ciencia biológica – cuya autoridad al respecto se contemplaba en principio como

³³⁹ Okasha en (2002, p. 196) es muy explícito al respecto: “Virtualmente todos los filósofos de la biología están de acuerdo en que [...] las especies no están individuadas por caracteres esenciales [...] Desde un punto de vista empírico, sencillamente no es verdad que los grupos de organismos que los biólogos tratan como co-específicos compartan un conjunto común de rasgos morfológicos, fisiológicos o genéticos que los diferencien de otras especies. Es cierto que para muchos grupos de organismos, el uso de una variedad de criterios de diagnóstico, normalmente basados en la morfología gruesa [“gross morphology”] de los organismos, permite que sean asignados a especies distintas de forma relativamente no ambigua”.

determinante, según recoge (b) – no clasifica a las entidades en géneros en función de la posesión de una misma estructura subyacente compartida, según reza (c). Ante esta tesitura, un defensor de (2) tiene dos alternativas. En primer lugar, podría concluir que los términos que designan especies biológicas no son términos de género natural, oponiéndose con ello a la concepción más usual de estos términos.³⁴⁰ En segundo lugar, podría debilitar lo establecido en (b) y sostener que, en ocasiones, puede obviarse el dictamen de la ciencia acerca de la elaboración de clasificaciones de seres naturales. La dificultad que ha de afrontar quienquiera que adopte esta segunda posición es la siguiente. Por una parte, su teoría de los géneros naturales se sustenta en el principio de que la ciencia posee el papel de instancia determinante para establecer una taxonomía de los géneros naturales, ya que, según se decía, las clasificaciones ordinarias tienen una validez supeditada a los hallazgos científicos. Pero, por otra parte, ante el hecho de que la propia ciencia y, en particular, la biología, admite una variedad de criterios en la determinación de las especies – como consecuencia de lo cual dichas especies poseen en ocasiones una composición heterogénea –, el partidario de esta teoría de los géneros naturales opta por despojar a la ciencia de ese papel determinante que le otorgaba de partida. Con otras palabras, se concede autoridad a la ciencia, pero a condición de que las taxonomías producidas en su seno no contravengan aquello que, de acuerdo con esta teoría de los géneros naturales, la ciencia ha de hacer, a saber, desarrollar las clasificaciones de los seres naturales a partir de una única propiedad subyacente compartida entre sus ejemplares.

En segundo lugar, nos oponemos a una actitud eliminativista con respecto a los términos de género natural introducidos en contextos ordinarios porque, a pesar de que estas categorizaciones pueden responder a intereses muy variados y no meramente científicos, ello no significa que las diferencias en las entidades de la naturaleza en virtud de las cuales se introducen tales términos carezcan en absoluto de importancia. De este modo, si nuestro trato con las entidades que clasificamos como repollos y el trato que reservamos a las entidades que denominamos “coles de Bruselas” es un trato distinto en muchos aspectos (procedimiento de cultivo, elaboración culinaria, consecuencias fisiológicas de su consumo, etc.); si, además, estas clasificaciones

³⁴⁰ Ésta es la posición de Ellis en (2001) y (2002). De acuerdo con este autor, las sustancias físicas y químicas, así como los átomos, moléculas y partículas subatómicas son géneros naturales. Sin embargo, la misma exigencia de que los ejemplares que pertenecen a un mismo género natural deben poseer una naturaleza intrínseca homogénea y unos poderes causales fijos y estables impide que las especies biológicas puedan ser consideradas géneros naturales.

resultan perfectamente útiles para nuestras inferencias y responden a diferencias reales – macroscópicas o no – en las entidades de la naturaleza que clasificamos en una u otra categoría, creemos estar perfectamente legitimados para considerar que los términos relevantes son términos de género natural. Aunque somos conscientes de que se trata de una postura controvertida, en este caso nos encontramos cercanos a la actitud de tolerancia hacia las taxonomías ordinarias defendida en el marco del realismo promiscuo de J. Dupré, según el cual las distinciones entre los géneros naturales son perfectamente reales y objetivas, pero esto no implica que haya una única clasificación legítima de los géneros naturales. Este autor propone, en consecuencia, una *noción laxa de género natural* de acuerdo con la cual los géneros naturales son formas de clasificar las entidades que toman en consideración propiedades objetivas de los mismos, en función de unos propósitos de investigación suficientemente bien articulados (Dupré 2002). Esta noción de género natural encaja bien con uno de los aspectos de nuestra posición acerca del fenómeno de la división del trabajo lingüístico, a saber, que la comunidad científica no es el único subconjunto de hablantes que es objeto de deferencia, debido a que otras actividades humanas distintas de la ciencia pueden plantear sus propios propósitos de investigación.

A modo de recapitulación, recogeremos aquellas razones que nos han permitido debilitar el supuesto (2) de la teoría de los géneros naturales que sustenta el externismo estrictamente natural, supuesto consistente en que tanto en los usos ordinarios como en los usos pre-científicos de los términos de género natural se presupone la existencia de una naturaleza compartida entre los ejemplares del género revelada por la ciencia (o susceptible de serlo). Así, en primer lugar, el hecho de que los usos ordinarios de esta clase de términos (tanto pre-científicos como posteriores al descubrimiento de esas presuntas propiedades subyacentes) no se rijan por las mismas exigencias que los usos científicos de estas expresiones hace cuestionable que las intenciones referenciales que Putnam atribuye a los hablantes ordinarios permitan defender la existencia de una continuidad sustancial entre estas dos clases de usos. En segundo lugar, se ha defendido que existen diferencias importantes entre los criterios efectivos de la propia ciencia a la hora de determinar los géneros naturales (en especial esos géneros que son las especies biológicas) y los criterios que, según esta teoría de los géneros naturales, rigen la práctica taxonómica de las ciencias. En tercer lugar, las taxonomías ordinarias responden a sus propios criterios y exigencias, aun cuando éstos sean distintos de los

científicos, determinados por la clase de trato que los hablantes quieran establecer con las sustancias y otras entidades naturales.

Llegado a este punto, nos ocuparemos del supuesto (3) que sustenta el externismo estrictamente natural, el supuesto de que las propiedades subyacentes de los miembros de los géneros naturales son descubiertas y no estipuladas. La validez general de este supuesto, en primer lugar, ha sido cuestionada a partir del examen de casos reales que muestran que, en relación a cierto tipo de géneros naturales, hay un componente estipulativo. A este respecto Beebe y Sabbarton-Leary (2010a, pp. 164 y ss.) han reparado en la existencia de cierto tipo de términos de género natural que designan sustancias químicas, a los que denominan *descriptores*,³⁴¹ cuya referencia se ha fijado a partir de la estipulación (atributiva) de que cualesquiera entidades que posean una determinada propiedad formarán parte de la extensión del género relevante. Por ejemplo, el término “ununbio” se introdujo a finales de los años 90 para designar el elemento químico de número atómico 112.³⁴² Aunque en el momento en que se introdujo el término no existían pruebas concluyentes de que tuviera una extensión efectiva, dicho elemento ha terminado por añadirse a la tabla periódica tras obtenerse las evidencias requeridas para establecer su existencia. Lo interesante de los descriptores es que la relación entre la introducción del término y el posterior descubrimiento de la “esencia” del género natural es *exactamente la inversa* que la señalada por Putnam. Para este autor, primero se introduce un término para designar un género natural y, posteriormente, la comunidad científica *descubre* cuál es la esencia del género, la cual, a su vez, constituye la condición necesaria y suficiente de pertenencia al mismo. Sin embargo, Beebe y Sabbarton-Leary ofrecen un ejemplo real en el que no tiene sentido hablar de un descubrimiento de propiedades subyacentes del género posterior a la introducción del término, sino que la introducción misma del término tiene lugar a partir de una estipulación del tipo “denominaremos ‘ununbio’ a aquel elemento químico de número atómico 112”. Según se desprende de estas consideraciones, es posible encontrar ejemplos de términos de género natural cuya identificación con la propiedad

³⁴¹ Obviamente, el término “descriptor” empleado por estos autores tiene un sentido distinto al usual en lógica: el operador con cuya ayuda se forman las descripciones.

³⁴² El nombre de este elemento ha sido construido a partir de la descripción de su número atómico: las dos primeras sílabas, “unun”, indican los dos primeros numerales (“1”) del número atómico y la siguiente sílaba, “bi”, se corresponde con el último numeral (“2”).

que especifica la presunta esencia del género no tiene lugar *a posteriori*, como Putnam defiende.³⁴³

Ahora bien, un partidario de la concepción de los géneros naturales que subyace al externismo estrictamente natural podría aceptar que los descriptores constituyen una excepción a su teoría, la cual no tendría por qué verse afectada en relación a los demás términos de género natural. Sin embargo, existen evidencias de que la existencia de un componente estipulativo en la determinación de la referencia de otros términos de género natural distintos de los descriptores es un fenómeno ubicuo. Así, se ha señalado que cuando se produce un descubrimiento científico en relación a algún género natural K que era designado en contextos ordinarios mediante el término “K” hay un cierto *margen de decisión* en la comunidad científica que lleva a cabo el descubrimiento con respecto a la posibilidad de imponer revisiones sobre tales usos ordinarios de ese término “K”. En este sentido, LaPorte mantiene que en el caso del género natural ballena, los expertos descubrieron que estos animales eran mamíferos, dado lo cual optaron por corregir la creencia (anterior al descubrimiento) de que estos animales, las ballenas, son peces. La cuestión es que, según este autor, el establecimiento de esta conclusión no tiene nada de irresistible, ya que los científicos bien podrían haber tomado una decisión distinta, como la de ampliar la extensión del término “pez”, de tal manera que el enunciado “las ballenas son peces” siguiese siendo verdadero.³⁴⁴ Asimismo, LaPorte nos proporciona un ejemplo muy relevante para el caso que nos ocupa, a saber, el relativo al término “jade”.³⁴⁵ De acuerdo con él,³⁴⁶ los hablantes pasaron a designar mediante el término “jade” una sustancia, la jadeita, cuya composición química es distinta de la composición química de la sustancia a la que

³⁴³ Una consecuencia que se sigue de esta observación es que la verdad del enunciado “si x es ununbio, entonces x tiene 112 de número atómico” no es revisable (en el peor de los casos podríamos descubrir que no hay ningún ejemplar de ununbio), mientras que para Putnam la especificación de la propiedad esencial de un género natural que obtiene la investigación científica es perfectamente revisable a la luz de nuevos resultados.

³⁴⁴ LaPorte (2010), pp. 105 y ss. y (2004), pp. 63 y ss.

³⁴⁵ Desde luego, un partidario de la teoría de los géneros naturales que está siendo objeto de crítica podría negar que el término “jade” sea un término de género natural. En este sentido Bird discute en (2010) los ejemplos que LaPorte utiliza para alegar que el estatus de un término lingüístico en tanto que término de género natural es algo que depende de nuestras decisiones (históricamente situadas y perfectamente contingentes). La idea fundamental de Bird es que el hecho de usar un mismo término, “jade”, para designar entidades que no comparten una naturaleza intrínseca es algo que depende, en efecto, de nuestras decisiones relativas a intereses, pero se trata siempre de decisiones extra-científicas que no tienen nada que decir a la hora de determinar qué términos son términos de género natural. Para Bird, sólo deberíamos tener en cuenta el punto de vista de la química a la hora de establecer si un término que se presume de género natural designa una sustancia dotada de homogeneidad intrínseca y, por tanto, se trata de un genuino término de género natural.

³⁴⁶ Cf. LaPorte (2004), pp. 94 y ss., así como LaPorte (2010), pp. 106 y ss.

anteriormente habían designado mediante ese término, la nefrita.³⁴⁷ Ahora bien, el descubrimiento de la diferente composición química de esas sustancias, denominadas igualmente “jade”, no trajo consigo un cambio en los usos del término, sino que los hablantes y, en particular, los expertos chinos que elaboraban joyas a partir de ambos materiales, optaron por considerar que las porciones de ambos minerales pertenecen igualmente a su extensión. Si estas consideraciones son correctas, parece que la lectura externista del escenario de la Tierra Gemela de Putnam no es la única posible. Puesto que el núcleo de la misma consiste en que ante el descubrimiento de un líquido, el XYZ, superficialmente idéntico al agua pero con una composición química diferente, sólo cabe declarar que este líquido no es agua, ahora debemos enfrentarnos a un ejemplo real de dos sustancias con distinta composición química que han recibido de hecho la misma denominación.

Por todo ello, es plausible sostener que los hablantes, incluidos los expertos, poseen un cierto poder de decisión con respecto a qué factores son relevantes para aplicar un término de género natural, de acuerdo con lo cual o bien pueden optar por excluir ciertos ejemplares de la extensión de un término si descubren que existe una diferencia relevante entre éstos y otros ejemplares que seguirán perteneciendo a la extensión, o bien pueden tomar la decisión de que la diferencia descubierta es determinante para modificar la extensión del término. En cualquier caso, estas consideraciones debilitan la tesis de que las esencias subyacentes, al ser descubiertas por los científicos, impondrían *de suyo* un modo de clasificar los géneros naturales.³⁴⁸

En cuanto al supuesto (4) que sustenta las tesis del externismo estrictamente natural, es decir, el supuesto de que los hablantes profanos están dispuestos a rectificar su uso de los términos de género natural en función del dictamen de los expertos sobre la posesión (o no) de una misma estructura subyacente por parte de un conjunto de entidades, tampoco se trata de una idea exenta de polémica. A este respecto, hemos argumentado en relación al supuesto (2) que ni siquiera está claro que los hablantes

³⁴⁷ Vid. nota 201.

³⁴⁸ Existe otro ejemplo, mencionado por Dupré en (1993, pp. 25-26) y (2002, pp. 23-24), en el cual el hecho de descubrir que existen diferencias significativas entre dos conjuntos de entidades que hasta ese momento se denominaban haciendo uso del mismo término “T” no supone un cambio en los usos ordinarios de la expresión ni conlleva negar que alguno de los dos grupos de entidades son T, algo que era fundamental en el experimento mental de la Tierra Gemela en relación a la sustancia cuya composición es XYZ. Este autor describe el caso de las hayas europeas y norteamericanas que, aunque en las taxonomías de la botánica son consideradas distintas especies, se denominan haciendo uso del mismo término de género natural. La moraleja que extrae Dupré a partir de ello es que un descubrimiento científico de este tipo no siempre tiene los efectos que Putnam pretende.

ordinarios (o aquellos que usaban un determinado término de género natural antes del descubrimiento de la presunta estructura subyacente que determinaría la extensión de dicho término) hayan tenido la intención de referirse a entidades dotadas de una misma “estructura oculta común”, por lo que los efectos de un presunto descubrimiento de que tal estructura existe (o no) en tales entidades adolecen de un grado de indeterminación considerable.

Por otro lado, la actitud deferencial hacia aquellos hablantes que podrían determinar, según se pretende, las propiedades subyacentes de los géneros naturales en un tiempo futuro no resulta un rasgo exportable a *todos* los hablantes de *cualquier* lenguaje natural, sino que, más bien, parece restringirse a aquellos lenguajes insertos en nuestra cultura, la cual concede una importancia muy destacada a la ciencia.³⁴⁹ Este argumento menoscaba la legitimidad que los partidarios del externismo estrictamente natural pretenden poseer para determinar qué concepto expresaban los hablantes del pasado mediante el uso de un término de género natural. Según vimos, esta clase de externistas consideran que el descubrimiento de la estructura subyacente del género impone una norma para la aplicación del término relevante (“sólo si algo tiene la propiedad P debe denominarse T”), norma que puede ser utilizada para examinar casos anteriores al descubrimiento de la propiedad (“los hablantes del pasado tenían la intención de referirse a entidades dotadas de una naturaleza común mediante el uso del término ‘T’”). El problema, sin embargo, es que estas tesis han visto reducida su plausibilidad por todos los argumentos que hemos enunciado para cuestionar la existencia tanto de este tipo de propiedades como de una actitud deferencial hacia los hablantes encargados de determinar las supuestas propiedades subyacentes, así como por aquellos contraejemplos que muestran que, en ocasiones, se sabe que existe una diferencia en las propiedades subyacentes de un conjunto de ejemplares denominados por un mismo término de género natural sin que ello se traduzca en una diferencia de uso (por ejemplo, el término “jade”). Por este motivo, pensamos que, frente a lo mantenido por los partidarios del externismo estrictamente natural, no hay una forma clara y directa de acercarnos a esos casos previos al conocimiento de las propiedades subyacentes, ya que, aún suponiendo que ahora sabemos cuál es la estructura

³⁴⁹ En este sentido Beattie (1993, pp. 321 y ss.) califica de “provinciana” a la propuesta de Putnam. A este respecto cabe decir que para muchas culturas son los ancestros o los dioses locales quienes dieron nombre a las cosas en el principio de los tiempos. Por tanto, en su caso la aplicación de un término será correcta en la medida en que se adecue al estándar instituido por esos seres a los que se les profesa reverencia más bien que por unos criterios aún por venir (Ardévol/Munilla/Cervelló 2003, pp. 110 y ss.).

subyacente de muchas sustancias (lo cual ya es mucho suponer), hemos comprobado que las normas para usar aquellos términos que las designan no siempre se han visto afectadas por ese conocimiento. El examen del comportamiento lingüístico de los usuarios de términos de género natural que vivieron antes de esta clase de descubrimientos arroja, en consecuencia, unos resultados más débiles de lo que inicialmente se pretende.

En resumidas cuentas, hemos tomado en consideración aquellos argumentos que, en nuestra opinión, tienen una mayor efectividad contra la concepción de los géneros naturales y de la referencia de los términos de género natural que mantienen los partidarios del externismo estrictamente natural. Desde luego, se ha aclarado desde el comienzo de la presente subsección que el fin perseguido no ha sido en ningún caso agotar el repertorio de problemas relacionados con el ejemplo particular del agua y con la teoría de los géneros naturales y de la referencia de los términos de género natural mencionada. Nuestra pretensión era mostrar que la lectura externista del primer escenario que describe Putnam en (1975c), lectura en la que se basa el externismo estrictamente natural, está lejos de ser el único acercamiento posible o, simplemente, el más intuitivo a la situación descrita, ya que requiere suscribir una serie de supuestos considerablemente polémicos acerca de los géneros naturales y acerca de la determinación de la referencia de los términos de género natural. De ser como decimos, cualquier autor que esté dispuesto a defender la concepción estrictamente natural del externismo habrá de asumir, ya desde el principio, unos supuestos muy discutibles.

4.4 *Influencia natural directa y atribuciones mentales. Nuestra propuesta: influencia natural mediada no-esencialista*

En lo que sigue intentaremos mostrar que el externismo estrictamente natural no sólo queda lastrado por los problemas descritos sino que, además, constituye una versión del externismo psicológico que no aporta ventajas significativas en la explicación de las prácticas de atribución de contenidos mentales, ya que, como hemos adelantado, resulta especialmente agresiva con el acceso privilegiado.

Por lo que se refiere a lo primero, recordemos que esta variante de externismo sitúa entre las condiciones de individuación de nuestros conceptos ciertos factores (i) que el hablante puede no conocer, (ii) que la comunidad en la que está inserto el hablante puede no conocer tampoco y que, en último término, (iii) son factores que

pueden no llegar a ser conocidos jamás. En relación al rasgo (i), se trata de una tesis presente en todas las variantes de externismo, a saber, la tesis de que deben incluirse entre las condiciones de determinación de los contenidos psicológicos de un hablante factores que no son internos a su mente individual. Sin embargo, ya explicamos que los dos últimos supuestos son exclusivos del externismo estrictamente natural, así como del externismo de distribución que integra las conclusiones del primero en cuanto a la individuación de los conceptos de género natural. En lo que concierne al rasgo (ii), el propio experimento mental de la Tierra Gemela en torno al uso del término “agua” establece diferencias en los contenidos mentales de dos hablantes basándose en factores que les resultan desconocidos a ellos mismos y a cualquier otro hablante de la comunidad a la que pertenecen. Ahora bien, llegado este punto es razonable preguntarse qué clase de relevancia podría tener una distinción conceptual como ésta. O dicho de otra manera, parece que el hecho de pensar, como haría cualquier habitante de la Tierra y de la Tierra Gemela en el tiempo en el que se desarrolla el experimento,³⁵⁰ que ambas sustancias son exactamente la misma, que no hay diferencia en sus usos ni en las funciones en las que se ven involucradas, conlleva que establecer una diferencia en el concepto que los habitantes de un entorno y otro expresan haciendo uso del término “agua” es establecer una distinción perfectamente prescindible para explicar sus actos de comunicación ordinarios. En este sentido, creemos que se trataría de una distinción que no juega ningún papel en la práctica lingüística, que no ayuda a resolver ningún malentendido en ningún intercambio comunicativo entre los hablantes de estas comunidades y que no sirve para explicar mejor el comportamiento lingüístico y no lingüístico de éstos. En palabras de Marconi, “[...] antes de 1750, y siguiendo a Putnam, en la Tierra Gemela estarían comunicándose con éxito, coherente y consensualmente acerca de lo que ellos consideran el mundo real. Puede que en esto [en el hecho de creer que ambas sustancias son idénticas] se equivocaran, pero su error no afectaría a sus intenciones – pretendían comunicarse sobre el mundo real y el agua de verdad, no solamente transmitir ‘los contenidos solipsistas de sus mentes’ – ni tampoco al éxito de sus actos de comunicación, dado que se entenderían perfectamente unos a otros” (1997, p. 111).

³⁵⁰ Según hemos podido ver, existen buenas razones para pensar que *incluso hoy en día* el XYZ podría considerarse un tipo de agua (puesto que hoy reconocemos la existencia de varias subclases de agua) o, si no se quiere ir tan lejos, para negar que el XYZ sea agua pero reconocer, al mismo tiempo, que este hecho depende más de contingencias históricas en torno a la estipulación de un uso para el término “agua” que de descubrimientos de supuestas esencias.

Asimismo, tal y como se vio en el capítulo dedicado a los casos de traslado, un hablante de la Tierra que sufre un transporte inadvertido a la Tierra Gemela podría permanecer ignorante de este hecho toda su vida. Probablemente establecerá enseguida múltiples contactos con el líquido relevante de ese entorno (lo beberá, se bañará en él, se resguardará de la lluvia, etc.), debido a lo cual, según el partidario del externismo estrictamente natural, el concepto que expresa mediante el término “agua” terminará por cambiar, aunque esto ocurrirá sin que él lo advierta y sin que nadie en su entorno lo advierta tampoco. Las consecuencias de esto son, en el peor de los casos, que el hablante sufrirá un cambio conceptual sin saberlo,³⁵¹ que este cambio hará falsas una serie de creencias sobre hechos previos al traslado³⁵² y que ello propiciará que cometa falacias de equivocidad en aquellos argumentos en los que una de las premisas expresa una creencia de esa clase.³⁵³ Pero *nada más* (o nada menos). Lo que queremos decir es que la clase de resultados derivados de ese cambio conceptual con el que se ve comprometido el partidario del externismo estrictamente natural no tienen nada que ver con dificultades en la relación lingüística y no lingüística de un hablante con sus interlocutores, ni le impedirá tener éxito en su trato práctico con ambas sustancias y saciar su sed al beber un vaso de lo que denomina “agua” en cada entorno. En suma, pensamos que establecer una diferencia entre el concepto que este hablante expresaba antes haciendo uso del término “agua” y el concepto que expresa ahora, tras el traslado, supone respaldar una teoría de la individuación conceptual basada en supuestos semánticos y metafísicos cuestionables, que está comprometida con cambios conceptuales inadvertidos y que, además, no nos ayuda en absoluto a entender mejor la conducta (lingüística o no) de los hablantes, lo cual debería constituir, en nuestra opinión, el objetivo fundamental de las atribuciones mentales, propias y ajenas.

Por otra parte, en lo que se refiere al tipo (iii) de condiciones de individuación que el externismo estrictamente natural incorpora, es decir, condiciones de individuación que podrían llegar a no conocerse nunca, las consecuencias que se derivan de su integración son, aún si cabe, más inverosímiles. De nuevo, acudamos al ejemplo de Liu (2002) sobre el “error de los expertos” para comprobar que, en efecto, en tanto que el externismo estrictamente natural determina la identidad/diferencia de

³⁵¹ En el tercer capítulo de nuestro trabajo comprobamos que muchos autores compatibilistas han optado por argumentar que los cambios conceptuales no suponen un atentado contra el acceso privilegiado en la medida en que éste no incluye la facultad de tener “conocimiento comparativo de nuestros conceptos”. Cf. Burge (2003a), Falvey y Owens (1994), Goldberg (2003) y Bernecker (1997).

³⁵² Si aceptamos la argumentación de Boghossian en (1989) y (1992).

³⁵³ Cf. Boghossian (1994).

nuestros conceptos en función de “esencias”, “estructuras ocultas” y “propiedades subyacentes” de los géneros naturales (cuyo conocimiento podría no obtenerse nunca), muchas de nuestras atribuciones conceptuales que conciernen a los pensamientos sobre géneros naturales podrían permanecer por siempre equivocadas sin que nadie llegara a saberlo jamás. Es decir, dado que cabe la posibilidad de que no lleguemos a acceder a las genuinas condiciones de individuación de los conceptos de género natural, podría ocurrir que no individuásemos correctamente ninguno de estos conceptos. En este sentido, de la misma manera que Putnam describe una Tierra Gemela en la que la única diferencia con respecto a la Tierra es una diferencia que permanece oculta (durante un tiempo) a los hablantes de ambos entornos, nada impide concebir otro entorno contrafáctico en el que cada género natural del mundo actual tiene una réplica macroscópicamente idéntica en ese mundo pero cuyas propiedades subyacentes determinan distintos géneros naturales.³⁵⁴ Si dichas propiedades nunca hubieran llegado a ser descubiertas, se seguiría de ello que ningún hablante de ninguno de estos dos entornos habría podido distinguir el género natural del entorno actual de su réplica contrafáctica (esencialmente distinta) y, sin embargo, para el partidario del externismo estrictamente natural todos los conceptos relevantes habrían sido diferentes.

A nuestro modo de ver, una teoría como ésta resulta extremadamente revisionista en el ámbito de nuestros hábitos de atribución mental, máxime cuando, tal y como acabamos de comprobar, dichos hábitos han funcionado con un éxito considerable. De manera similar a la tesis que Wikforss defiende en (2010, p. 165), según la cual deberíamos ser capaces de elaborar nuestra semántica con independencia de la ciencia y la metafísica, nosotros abogamos por una teoría de la individuación conceptual que sí tome en consideración logros importantes de nuestra ciencia y aquellos supuestos metafísicos más acordes con tales logros, pero que no se convierta en esclava de una metafísica esencialista o de unos conocimientos científicos que podrían no alcanzarse nunca. Si nuestras observaciones son correctas, parece recomendable rechazar las tesis no sólo del externismo estrictamente natural sino también de aquellas versiones del externismo psicológico que incorporan dichas tesis (el externismo de distribución).

En lo sucesivo tomaremos en consideración la última variante del externismo, el externismo social, en cuyo seno, como indicamos, cabe distinguir dos versiones. Por un lado, tenemos una concepción esencialista, que individúa los conceptos de género

³⁵⁴ Es decir, nada impide concebir un escenario de estas características siempre y cuando se acepte que esta clase de escenarios tienen relevancia para aclarar y probar nuestras intuiciones semánticas.

natural en función de las propiedades subyacentes de los géneros naturales, siempre que el conocimiento de esas supuestas propiedades forme parte del patrimonio del conocimiento social; por otro lado, encontramos una versión no-esencialista del externismo social, de acuerdo con la cual los conceptos de género natural son individuados en función de los usos comunitariamente asignados a estos términos, sin necesidad de que tales usos hagan alusión a propiedades subyacentes, tal y como los partidarios del externismo estrictamente natural entienden dichas propiedades. Nuestra siguiente tarea consistirá, en primer lugar, en justificar que existen ciertas tensiones internas a la concepción esencialista del externismo social y, en segundo lugar, en abogar por una visión no-esencialista del mismo, mostrando que la principal objeción contra esta teoría – que no reserva un tratamiento especial a los conceptos de género natural – no constituye un problema insuperable.

Con respecto al primer objetivo, pensamos que la noción misma de estructura subyacente manejada por los partidarios del externismo estrictamente natural presenta ciertas fricciones al ser incorporada al marco del externismo social, aunque éste se conciba de forma esencialista. Así, en primer lugar, reparemos en el hecho de que al hablar de propiedades subyacentes estamos refiriéndonos, por lo general, a propiedades que, por un lado, son distintas de las propiedades macroscópicas o superficiales y, por otro, pretenden servir para determinar de manera unívoca qué entidades pertenecen a la extensión de un término de género natural. En lo que concierne al primer rasgo, es de suponer que en relación a multitud de lenguajes naturales, la mayor parte de las convenciones asociadas socialmente a los términos de género natural tienen que ver con los aspectos más reconocibles de las instancias que pertenecen al género, más bien que con sus propiedades subyacentes. La razón es que, por una parte, hace falta un gran trabajo científico para detectar la clase de propiedades que suelen considerarse propiedades subyacentes paradigmáticas (composición química, secuencia de ADN, etc.) y, sin embargo, antes de alcanzar un conocimiento tan sofisticado, los hablantes usaban de manera reglada términos de género natural y expresaban pensamientos mediante el uso de oraciones en las que figuraban tales términos. Por otra parte, si hay algún aspecto del concepto expresado por esta clase de términos que es especialmente susceptible de ser desconocido o ser comprendido sólo en parte es, precisamente, el que concierne a las propiedades subyacentes. Por estas razones, parece que aunque los estándares que sancionan el uso de los términos de género natural pudieran recoger la clase de propiedades que se consideran esenciales *cuando se ha alcanzado un*

desarrollo científico suficiente, también harán alusión, en gran medida, a la posesión de ciertas propiedades superficiales por parte de una entidad como condición de aplicación del término.

Teniendo esto en cuenta, vamos a examinar qué diría un partidario del externismo social esencialista acerca de la identidad/diferencia de los conceptos relevantes en el caso presentado en el experimento mental de la Tierra Gemela en torno al uso del término “agua”. Parece admisible que *siempre y cuando* la diferencia en los líquidos haya sido descubierta, el partidario de esta comprensión del externismo está autorizado a establecer una diferencia en el concepto que los hablantes de ambos entornos expresan haciendo uso del término “agua”. Sin embargo, sería interesante preguntarnos por su diagnóstico en cuanto a la identidad/diferencia del concepto en cuestión *antes* de que el conocimiento de la estructura subyacente haya sido incorporado a las condiciones socialmente determinadas para la aplicación del término. A este respecto, por una parte, parece claro que su teoría no invita a establecer ninguna diferencia en el concepto que los hablantes expresaban en tal secuencia temporal haciendo uso del término relevante. En cambio, dado que el externismo social esencialista incorpora la noción de esencia que pone en juego el partidario del externismo estrictamente natural, dicha noción empuja nuestras intuiciones en la dirección opuesta y requiere establecer una diferencia en el concepto en cuestión. Para comprobarlo, recordemos que, según la mayor parte de los defensores de esta última teoría, las propiedades subyacentes son objeto de descubrimiento (más bien que de estipulación) y, además, se trata de propiedades que determinan la pertenencia al género. Por este motivo, de admitir la noción de esencia que maneja el externismo estrictamente natural, resulta razonable juzgar el uso de los términos (y la individuación de los conceptos correspondientes) a la luz de los descubrimientos de las propiedades subyacentes. Esto es, si aceptamos que *hemos descubierto* que K es P (siendo K un género natural y P la estructura subyacente que lo constituye) y que cualquier otra cosa distinta de K no es P, entonces resulta razonable extraer de dicho descubrimiento una norma sobre el uso del término “K”, la cual, además, tendrá un poder de sanción normativa que se extiende al tiempo anterior al descubrimiento (“nos equivocamos al llamar “K” a esa clase de entidades que no eran P”). En este sentido, un partidario de la clase de teoría esencialista de los géneros naturales que sustenta el externismo estrictamente natural nos diría que al *descubrir* que un mismo término se ha aplicado, en realidad, a ejemplares *esencialmente* distintos, deberíamos efectuar correcciones

sobre los usos pasados del término relevante (y pensar, así, cosas como “erramos al creer que los hablantes de cada entorno pensaban en lo mismo cuando hacían uso del término”, “en realidad no expresaban el mismo concepto a pesar de usar el mismo término y de ser incapaces de diferenciar aquello a lo que se estaban refiriendo”, etc.). Ahora bien, debemos preguntarnos si un defensor de la individuación social de los contenidos mentales, que considera que los usos convencionalmente asociados a los términos son *los únicos factores externos* que determinan nuestros conceptos, tiene derecho a imponer tal revisión, puesto que, tal y como vimos, antes del descubrimiento de la supuesta diferencia esencial entre los dos conjuntos de ejemplares denominados “K”, los usos asociados a ese término eran, en ambos casos, exactamente los mismos.

Según podemos apreciar, la concepción esencialista de los géneros naturales mencionada conlleva que el presunto descubrimiento (siempre revisable, eso sí) de la estructura subyacente de un género natural revela la norma fundamental de la aplicación del término, una norma que puede haber permanecido ignorada por los usuarios del lenguaje durante siglos. Sin embargo, resulta un tanto inverosímil que una teoría de la individuación social de los contenidos reserve a esta norma la capacidad de proyectar un diagnóstico sobre la individuación de los conceptos que los hablantes expresaban haciendo uso del término de género natural en aquellos momentos en que era una norma desconocida para todos y cuyo poder de sanción era igual a cero. Una norma que nadie respeta porque nadie conoce no es, en puridad, una norma.³⁵⁵

En nuestra opinión, la opción más viable para dar cabida tanto a las intuiciones que respaldan a la mayor parte de experimentos que se presentan a favor del externismo psicológico³⁵⁶ como a aquellos fenómenos psico-lingüísticos que enumeramos en la segunda sección, es el externismo social no-esencialista. Esta opción contempla la deferencia como el factor de cohesión lingüística entre los hablantes de una comunidad, cuyo repertorio conceptual, por tanto, estaría sujeto al principio de parsimonia, en cuanto que no sería preciso multiplicar los conceptos expresados por los hablantes

³⁵⁵ En (1997, pp. 122- 123) Marconi se pronuncia en el mismo sentido: “Se puede objetar, a bote pronto, que las opiniones causal-externistas que conceden a los bautismos y otras ceremonias de denominación un papel esencial son, ciertamente, formas de normativismo, en las cuales la norma es fijada por el bautismo inicial [...] En lo sucesivo, si algo ha de ser llamado un olmo depende no de criterios individuales, ni siquiera comunitarios, sino sólo de si la norma es de aplicación, esto es, de si el objeto en cuestión tiene la misma naturaleza que los olmos paradigmáticos. Pero esto no funciona. Las normas tienen que ser aplicables: debe haber alguien que esté en condiciones de garantizar que se practica la norma”.

³⁵⁶ Esto es, a todos los experimentos mentales de Putnam y Burge, exceptuando el experimento mental en torno al uso del término “agua”.

atendiendo a la multitud de pensamientos idiosincrásicos que pueden mantener sobre los mismos. Si consideramos que los hablantes, por lo general, expresan mediante el uso de sus términos los conceptos que *se* expresan en su comunidad, podremos interpretar sus pensamientos tomando sus palabras en su valor aparente. En suma, abogamos por una visión externista de la individuación de nuestros contenidos mentales que conciba a los hablantes como usuarios de un lenguaje cuyas propiedades semánticas no están determinadas de manera individualista y que, al mismo tiempo, evite comprometerse con teorías de los géneros naturales que podrían lastrar su plausibilidad inicial.

No obstante, tal y como hemos señalado, nuestra posición debe hacer frente a la siguiente objeción. De considerar que la identidad de los conceptos de género natural no depende de aquellas propiedades subyacentes constitutivas de los géneros, parece que no habría ninguna diferencia entre la determinación de la identidad de esta clase de conceptos y la individuación de aquellos otros conceptos, distintos de los conceptos de género natural, susceptibles de recibir una individuación externista. O, dicho de otro modo, si optamos por rechazar la influencia natural directa y nuestra tesis sólo reconoce la influencia social y la influencia natural socialmente mediada en la determinación de la identidad de nuestros conceptos y contenidos, entonces estaríamos despojando a los conceptos expresados por términos de género natural de la especificidad que el externismo estrictamente natural les reservaba. En definitiva, este posible interlocutor alegraría que sólo podríamos conservar esa especificidad de los conceptos de género natural si, además de los factores sociales, se requieren factores de carácter natural para la individuación de los mismos.

Nuestra respuesta consiste, en primer lugar, en aceptar esta última alegación pero en rechazar, seguidamente, que la forma en la que nuestro interlocutor entiende la influencia de los factores naturales sea la única posible. Así, la tesis que defendemos no excluye la influencia de los factores naturales *in toto*, por lo cual podemos reservar un tratamiento distinto para los conceptos de género natural frente a otro tipo de conceptos. En este sentido, desde nuestro punto de vista sigue siendo perfectamente posible establecer una diferencia entre aquellos conceptos individuados puramente en virtud de consideraciones convencionales de carácter social (conceptos de género socio-legal, por ejemplo) y conceptos para cuya individuación hemos de atender a factores naturales que, aunque deben ser conocidos al menos por un grupo de expertos, no son factores que dependen del puro arbitrio de los hablantes. La justificación que podemos ofrecer al respecto es que la variante de externismo que mantenemos reconoce la existencia de

regularidades presentes en las entidades naturales que son empleadas como criterios de individuación y clasificación, aunque no necesariamente desde un punto de vista estrictamente científico. Cabe ofrecer un ejemplo en este sentido si pensamos en dos grupos de animales, A y B, que no tienen descendencia fértil cuando se cruzan entre sí. Esta regularidad podría ser tomada en consideración como un criterio para determinar que A y B constituyen especies biológicas distintas y, por tanto, como un indicio de que son géneros naturales diferentes. Sin embargo, no parece haber ningún factor natural que determine la diferencia entre el concepto expresado por el término “denuncia” y el expresado por el concepto “querella”, puesto que la distinción entre (los usos asociados a) un término y otro depende de la libre decisión de los hablantes de tomar como un rasgo relevante para distinguir entre ambos conceptos que el sujeto que pone en conocimiento del juez unos hechos que piensa constitutivos de delito sea parte acusadora en el proceso (querella) o no lo sea (denuncia). A partir de estos ejemplos se comprueba que los conceptos de género natural (frente a conceptos como el de denuncia o el de querella) sí serían “especiales” en el sentido de que ciertos factores naturales (las regularidades conocidas mencionadas) pueden figurar entre sus condiciones de individuación, aun cuando rechazamos la influencia de otro tipo de factores naturales, es decir, de aquellos cuyo poder de individuación se sitúa completamente al margen de las decisiones, intereses e intenciones lingüísticas de cualquier hablante.

Asimismo, podemos reforzar nuestra posición acudiendo a las observaciones de Rudd en (1997), quien defiende que la individuación externista en función de factores sociales se apoya en unas fuentes e intuiciones distintas de las que respaldan la individuación externista en función de factores naturales, entendidos del modo en que los entiende el partidario del externismo estrictamente natural. Precisamente, hemos defendido que esta última concepción de la individuación conceptual apela a una consideración de los géneros naturales sumamente cuestionable, pero, en el caso del externismo que defendemos, no se requiere apelar a esta clase de tesis metafísicas,³⁵⁷ sino sólo atender al hecho de que, por lo general, las normas públicas que determinan el uso de nuestros términos juegan un papel fundamental en la fijación de la identidad de los contenidos mentales. Según se ha visto, estos criterios no tienen por qué ser

³⁵⁷ Esto no significa que el partidario del externismo social no-esencialista rechace que no es posible encontrar una propiedad especialmente relevante para la individuación de un concepto de género natural. Lo que se rechaza son las implicaciones que tiene denominar “esencial” a dicha propiedad (oculta pero susceptible de ser descubierta – no estipulada –, única para todos los usos, determinada sólo por los científicos, proyectable hacia los usos pasados y futuros del término, etc.).

conocidos en su totalidad por el hablante individual, pero sí se requiere que éste se encuentre inserto en una red de relaciones sociales en la que existen grupos de hablantes o instituciones que establecen y revisan los criterios en cuestión. Asimismo, hemos señalado que no hay que entender las relaciones de deferencia de manera simplista, como si en cada caso existiera un único grupo de expertos relevantes en el uso del término, generalmente un grupo de científicos. En nuestro trato con las cosas y, en particular, con las clases de cosas (naturales o no) que conceptualizamos y designamos lingüísticamente podemos perseguir distintos fines e intereses, en razón de lo cual pueden existir diferentes grupos de expertos en el uso de un cierto término, con un patrimonio peculiar de conocimientos y habilidades en relación con las entidades conceptualizadas. Los hablantes pueden confiar en uno u otro en función de cuál sea el interés que dirija su acercamiento a los objetos en un momento determinado, puesto que “conservan el derecho a emplear los términos de distintas formas sujetas a la negociación social, pero en el seno de una sociedad concebida en un sentido amplio, dentro de la cual la comunidad científica tiene voz, pero es sólo una voz entre muchas” (Hampton 1998, p. 574). Esa es la razón, precisamente, de que en lo relativo a las entidades naturales respaldemos la aceptación de las taxonomías ordinarias que propugna J. Dupré, en la medida en que las ideas de este autor legitiman la introducción de términos de género natural que sirven para clasificar este tipo de entidades en función de intereses no científicos pero perfectamente relevantes en otros contextos.

En suma, si los factores internos al hablante en muchas ocasiones son insuficientes para fijar la identidad de sus conceptos ello se debe a que, por lo general, un hablante no conoce todos los criterios (públicos y revisables) para el uso de sus términos. Desde luego, suscribimos la opinión de Putnam cuando afirma que para determinar si un hablante posee un determinado concepto simplemente “observamos (y preguntamos, si es preciso) para comprobar si sabe cómo usar la palabra [correspondiente]” (Putnam 1975b, p. 8). Si este hablante usa la palabra de manera sintácticamente correcta, si es capaz de aplicar el término con un acierto considerable, si da respuestas mayoritariamente apropiadas cuando mantiene una conversación que involucra el término (aun cuando en dichas respuestas haya lagunas e incluso oraciones del tipo “desconozco tal aspecto de ese asunto”), entonces no se necesita más para determinar que ese sujeto posee el concepto relevante, a saber, aquél que *se* expresa en su comunidad lingüística haciendo uso de esa palabra.

No queríamos finalizar esta sección sin responder a una última objeción posible a nuestra tesis, según la cual el hecho de que exista más de un grupo de expertos en relación a un determinado concepto no encaja bien con la idea de que los hablantes de una comunidad lingüística, en tanto que tales, expresan un único concepto cuando hacen uso del término relevante y que éste, de acuerdo con el enfoque externista que defendemos, es un concepto distinto del que expresarían los hablantes de otra comunidad que establecieran un uso diferente para el término en cuestión. Según reza esta objeción, desde el momento en que hemos reconocido la existencia de varios subgrupos de hablantes expertos no cabría hablar de un único concepto, expresado por *todos* los hablantes de la comunidad en cuestión, sino de una multiplicidad de conceptos dentro de una misma comunidad, cuya identidad depende del subconjunto de hablantes en el que se delege en cada caso. Aunque dicha objeción presenta cierta plausibilidad inicial, no pensamos que el reconocimiento de múltiples grupos de deferencia aniquile en general la unidad conceptual comunitaria.

Para empezar, cabe señalar que la existencia de una multiplicidad de grupos de deferencia no es sólo una consecuencia de mantener una teoría de los géneros naturales tolerante hacia las taxonomías de sentido común, en cuyo caso se trataría de una tesis eliminable si se escoge otra teoría al respecto. Por el contrario, la multiplicidad de grupos de deferencia es un hecho que se pone de manifiesto en relación a varios fenómenos lingüísticos, tales como los mecanismos de creación de términos por parte de los miembros de estos subgrupos, mecanismos que en muchas ocasiones tienen como fin asegurar (a veces por razones sociológicas) su posición de *único* grupo de expertos relevante. De este modo, es frecuente que un determinado grupo de expertos acuñe un nuevo término para designar una realidad cuya denominación anterior está demasiado “extendida” entre los profanos o para impedir su apropiación por parte de otros grupos de expertos en competencia.³⁵⁸

Dicho esto, podemos argumentar que la existencia de múltiples subgrupos de expertos en una comunidad en relación a un mismo término no amenaza por lo general

³⁵⁸ La medicina es un campo del conocimiento en el que esta última clase de fenómenos se produce de manera frecuente, tal y como reconocen sus teóricos. Gracia/Lázaro se expresan así en (2009, pp. 21-22): “La medicina es una ciencia a la que le corresponde un cierto lenguaje. El lenguaje médico es muy amplio, tiene un vocabulario muy extenso y, a primera vista muy extraño y difícil de entender. Esto puede explicarse por la complejidad de los problemas médicos, que requieren un sistema de comunicación muy preciso. Pero la terminología médica tiene además una función protectora: la medicina es una actividad que siempre ha estado muy preocupada por conservar el secreto de sus actividades, y una forma ideal de conservar un secreto es hablar un idioma que los demás no entienden [...] Es frecuente que cuando un término médico pasa a ser conocido popularmente, los profesionales sanitarios dejen de usarlo”.

la unidad del concepto expresado por él acudiendo a las observaciones de Burge,³⁵⁹ según las cuales los miembros de un conjunto de expertos relevante pueden disputar sobre los criterios fundamentales de aplicación de *un mismo* concepto (Burge 1986c, pp. 256 y ss.), algo muy frecuente en los grupos de discusión formados por especialistas. Si esta clase de debates no atenta contra la unidad conceptual, tampoco tendría por qué ocurrir algo similar en caso de que reconociéramos la existencia de distintos grupos de expertos que debaten sobre las condiciones de aplicación de un concepto dentro de una misma comunidad.

Para reforzar esta tesis, vamos a examinar dos ejemplos. En primer lugar, puede constatarse que el hecho de la deferencia múltiple ha existido hasta hace poco³⁶⁰ en relación al concepto expresado por el término de género socio-legal “matrimonio”. Por una parte, la Real Academia de la Lengua Española, una institución de deferencia privilegiada para los castellanohablantes, establecía como primera acepción del término “unión de hombre y mujer concertada mediante determinados ritos o formalidades legales” y ninguna de las acepciones restantes contravenía el requisito de la diferencia sexual de los contrayentes. Ahora bien, es innegable reconocer también a los jueces como expertos en este ámbito puesto que, al fin y al cabo, juegan un papel imprescindible en la concertación de matrimonios, así como en su reconocimiento, impedimento y disolución. Sin embargo, los contrayentes de los matrimonios celebrados por este grupo de expertos pueden ser, indistintamente, del mismo o de distinto sexo. Por último, una cantidad considerable de castellanohablantes consideran que los sacerdotes católicos son expertos en matrimonios, en la medida en que les atribuyen el papel de ministros de tal sacramento. Sin embargo, desde su punto de vista, los matrimonios sólo pueden celebrarse entre personas de distinto sexo y, además, no pueden disolverse. El resultado es que durante todo este tiempo ha existido un conflicto entre, primeramente, las condiciones de aplicación del término “matrimonio” determinadas según los académicos de la lengua (requisito de diferencia sexual entre contrayentes, independientemente de su carácter disoluble); en segundo lugar, las condiciones de aplicación de ese término admitidas por los jueces (no hay requisito de diferencia sexual y reconocimiento de su carácter disoluble) y, en tercer y último lugar,

³⁵⁹ Naturalmente, pueden encontrarse casos en los que un subconjunto de hablantes toma la decisión de otorgar un nuevo significado a una determinada expresión lingüística. Este fenómeno, distinto del que discutimos, suele producirse en el contexto de creación de jergas.

³⁶⁰ Concretamente, hasta el 22 de Junio de 2012, cuando se decidió introducir una segunda acepción de este término que también permite denominar así a las uniones de parejas del mismo sexo.

las condiciones de aplicación de ese término según la Iglesia Católica (requisito de diferencia sexual y carácter indisoluble del matrimonio).

Asimismo, podemos pensar en un segundo ejemplo relacionado con el término “cáncer”, con respecto al cual existen distintos grupos de expertos dentro de una misma comunidad, uno de los cuales está formado por los médicos y otro, por terapeutas de Reiki, una clase de medicina alternativa. Poco importa que la autora o el lector no estén dispuestos a calificar de “expertos en cáncer” a estos últimos. Nuestro objetivo es ofrecer una *descripción* de lo que hacen los usuarios del término “cáncer” en nuestra comunidad lingüística y ocurre que parte de esos usuarios presenta una actitud de deferencia en cuestiones de salud con respecto a grupos de hablantes distintos de los médicos usuales.³⁶¹ En este caso, mientras que el médico “ortodoxo” ofrece un diagnóstico y un tratamiento para el cáncer de próstata, el terapeuta de Reiki considera que esa enfermedad está causada por un desajuste en la energía vital y será sanada cuando el chakra sacro (el centro de energía relevante) quede desbloqueado. Uno podría pensar, a partir de esto, que la deferencia en un grupo de expertos u otro determina una identidad distinta para el concepto expresado por el término “cáncer”. Ahora bien, repararemos en el hecho de que el partidario de las terapias alternativas ofrece, en efecto, explicaciones y tratamientos diferentes *para el cáncer de próstata*. Es decir, el maestro de Reiki no pone en duda que el paciente sufra *cáncer* y, en función de ello, ofrece una explicación distinta para esos síntomas que sufre y plantea una terapia distinta (y a veces complementaria) del tratamiento que propone el médico usual. De hecho, es perfectamente verosímil pensar en una discusión entre un médico y un partidario del Reiki sobre el potencial curativo de la imposición de manos para sanar el cáncer u otro tipo de enfermedades. Se opine lo que se opine sobre los argumentos de los contendientes, no hay duda de que se trata de un debate con cierto sentido, en el que

³⁶¹ En (2010) Rauti también indica que la mejor perspectiva para examinar el fenómeno de que, en ocasiones, los hablantes confían en distintos subgrupos *de expertos* que establecen normas para aplicación de un término no estrictamente coincidentes es una perspectiva de carácter no-epistémico, sino puramente descriptiva. Este enfoque tiene una cierta ventaja, a saber, que permite explicar los cauces efectivos de las relaciones de deferencia, en la medida en que no vincula la denominación de “expertos” a un grupo de hablantes en función del conocimiento que éstos tengan de cierta área de la realidad, sino que simplemente atiende al hecho de que existe un número considerable de individuos que confía en este grupo y que esta confianza tiene efectos en su comportamiento lingüístico. De este modo, si en el seno de una comunidad lingüística hay un grupo de hablantes que delega en otros el conocimiento de las condiciones de uso de términos de género sobrenatural como “hada”, “chakra” o “ángel”, entonces no hay impedimento en considerarlos como “expertos”, al menos desde el punto de vista de este grupo de individuos.

cabe ofrecer datos sobre la recuperación de pacientes con *esas* enfermedades o discutir sobre las causas (distales) de *estas* dolencias.

Ambos ejemplos nos sirven para comprobar que la deferencia en distintos grupos de expertos que ofrecen unos criterios de determinación distintos para un concepto no tiene por qué amenazar la unicidad del concepto cuyas condiciones de determinación se discuten. Precisamente, es la identidad del concepto compartido en una comunidad lo que hace posible articular discusiones entre los diferentes grupos de expertos que establecen condiciones de aplicación relativamente distintas para *un mismo concepto*.

5. Conclusiones

A lo largo de este último capítulo hemos emprendido una serie de tareas de carácter aclaratorio y argumentativo. En primer lugar, convinimos en la necesidad de establecer una clasificación de las distintas tesis que han recibido el nombre genérico de “externismo psicológico”. Aunque desde el primer momento declaramos que la variante de externismo cuya compatibilidad con el acceso privilegiado se cuestiona es la visión externista de la *individuación* de nuestros contenidos mentales, en algunas ocasiones se ha aplicado esa etiqueta a tesis que no tenían que ver estrictamente con la individuación de estos contenidos, sino más bien con su origen y especificación lingüística. Esto nos permitió presentar una taxonomía precisa y exhaustiva de los distintos aspectos de nuestros procesos y contenidos mentales (origen, especificación lingüística e individuación) que han sido examinados desde una perspectiva externista. En segundo lugar, expusimos la clasificación más usual de los diferentes modos de concebir la influencia de los factores externos en la individuación de los contenidos mentales y recogimos aquellos argumentos y fenómenos relacionados con el pensamiento y el lenguaje que, más allá de los experimentos mentales clásicos, aportan plausibilidad a las tesis externistas. No obstante, en esta segunda sección se defendió que esta tipología usual, que distingue entre la variante natural, la social y la mixta del externismo, presenta graves tensiones internas, especialmente en lo que se refiere a esta última concepción. Para demostrarlo, nos apoyamos en consideraciones de Burge en (1982), de Liu en (2002) e hicimos uso de ciertas conclusiones extraídas en los capítulos previos dedicados a los casos de traslado, con el propósito de poner en evidencia que la influencia de los factores de carácter natural, por un lado, y la influencia de los factores

de carácter social, por otro, determinan en ocasiones una identidad distinta para los contenidos mentales relevantes. Una vez concluida esta tarea, dedicamos la tercera sección a proponer una clasificación diferente de las variantes de externismo, con la exigencia de que cada una de dichas variantes nos permitiera obtener un diagnóstico coherente acerca de la individuación de nuestros estados mentales. Finalmente, la cuarta sección de este capítulo se reservó para el examen de aquellas tesis sobre los géneros naturales y sobre la semántica de los términos de género natural en las que se apoyan aquellas concepciones del externismo que dan cabida a la influencia de los factores naturales en la individuación de nuestros contenidos mentales, ya sea autorizando una influencia directa, no mediada socialmente, o ya se trate de una influencia que debe estar registrada en las convenciones lingüísticas que regulan el uso de los términos de género natural. En la última sección nuestra finalidad fue cuestionar la plausibilidad de estas dos concepciones del externismo a partir de dos líneas de argumentación. En primer lugar, defendimos que ambas concepciones se fundamentan en unas tesis sobre los géneros naturales y sobre la semántica de los términos de género natural muy controvertidas. En segundo lugar, mostramos que la clase de influencia del entorno que registran una y otra variante (sobre todo la primera) resulta especialmente agresiva con el acceso privilegiado a nuestros propios pensamientos, al establecer distinciones entre los mismos que, en el mejor de los casos, pueden resultar irrelevantes para las prácticas comunicativas de los usuarios del lenguaje y, en el peor, son distinciones cuyo fundamento *in rebus* podría no llegar a conocerse nunca.

En nuestra opinión, una concepción social del externismo, despojada de los compromisos esencialistas propios de las variantes naturales, constituye una teoría atenta a los criterios públicos de aplicación de nuestros términos y que permite explicar aspectos importantes de las prácticas ordinarias de atribución mental. La variante de externismo por la que abogamos se sustenta en una visión de los hablantes en tanto que agentes insertos en una comunidad lingüística que sanciona normativamente el uso de nuestros términos y cuya cohesión y solidez descansa en gran medida en actitudes de deferencia de unos hablantes hacia otros. Una teoría de la individuación mental como la que se propone nos permite explicar que, en la mayoría de los casos, lo que un hablante quiere decir con las palabras que usa en sus oraciones declarativas es lo mismo que un hablante de su comunidad diría cuando las usa. Sólo sobre la base de este acuerdo generalizado, revelado en la expresión “hablantes de *un* lenguaje (natural)”, es posible localizar y resolver los desacuerdos. Pensamos, en definitiva, que el externismo social

no-esencialista es una teoría útil y explicativa de hábitos efectivos de los hablantes en relación al fenómeno de las atribuciones mentales. Esta adecuación contrasta, en nuestra opinión, con las consecuencias revisionistas de aquellas variedades de externismo que se apoyan en una metafísica esencialista.

Capítulo 6: Conclusiones finales

A lo largo de este trabajo hemos tomado en consideración una serie de problemas que ha de afrontar la pretensión de compatibilizar el externismo psicológico con el acceso privilegiado a los propios pensamientos. Con el fin de precisar en qué punto reside la dificultad de esta tarea y qué argumentos se han desarrollado para poner de manifiesto dicha dificultad, dedicamos un primer capítulo a explicar los fundamentos semánticos de las tesis externistas en psicología. La justificación para proceder de esta manera consiste en que los orígenes de dichas tesis se sitúan, histórica y conceptualmente, en una reacción contra el enfoque clásico (descriptivista e internista) del análisis de las propiedades semánticas de dos clases de expresiones lingüísticas, a saber, los nombres propios y los términos de género natural. Asimismo, en este primer capítulo describimos con cierto detalle aquellos experimentos mentales que se han construido para aportar plausibilidad a la individuación externista de nuestros contenidos mentales, en especial, el experimento de Putnam en torno al uso del término “agua” y el experimento de Burge sobre el concepto de artritis. A partir de esto, alcanzamos un triple objetivo, a saber, (i) entender el externismo psicológico, en principio, como una extensión histórica del enfoque externista desarrollado en semántica, (ii) exponer los argumentos e intuiciones en los que encuentran apoyo los partidarios del externismo psicológico y (iii) obtener una comprensión lo suficientemente amplia del externismo como para permitirnos abordar el debate sobre su compatibilidad con el acceso privilegiado.

El segundo capítulo estuvo dedicado al análisis de los pasos iniciales de esta polémica, cuyos textos de referencia se publicaron en su mayor parte a lo largo de los años 80 del pasado siglo. Un rasgo digno de ser destacado en esta fase del debate es la ausencia casi completa de grandes textos incompatibilistas, un hecho que cambiará con la participación en los años posteriores de autores como Boghossian, McKinsey y Brown, cuyos trabajos inauguraron las etapas posteriores de la discusión. Sin embargo, la elaboración en los años 80 de argumentaciones a favor del compatibilismo (de la mano de Burge, Davidson y Heil) pone de manifiesto que por aquel entonces la comunidad filosófica había advertido la existencia de un problema, *prima facie*, para conjugar el externismo psicológico con el acceso privilegiado a los propios pensamientos. El “modelo basado en la inclusión”, que Burge planteó en estos primeros

trabajos, vino a dar respuesta a las incipientes inquietudes en este sentido. Por otra parte, el trilema desarrollado por Bilgrami en los primeros años de la década de los 90 constituye una excepción a este respecto, ya que ofrece un planteamiento incompatibilista considerablemente elaborado y de amplio alcance. No obstante, pudimos comprobar que el argumento de este autor presentaba un defecto fundamental, a saber, considerar que el externismo psicológico no requiere ningún tipo de mediación representacional en las relaciones cognitivas con el entorno natural y social, algo que, en rigor, no hace justicia a las propuestas de Putnam y Burge.

En lo que respecta a la segunda fase del debate, centrada en la discusión de los casos de traslado, el objetivo primordial de las intervenciones incompatibilistas es poner de manifiesto las insuficiencias del modelo basado en la inclusión diseñado por Burge. De este modo, aun cuando aceptáramos que dicho modelo consigue garantizar el acceso privilegiado a los pensamientos *presentes*, la posibilidad de ser trasladado de manera inadvertida a un entorno donde bien los géneros naturales, bien los usos lingüísticos socialmente establecidos, son diferentes sirve de fulcro para alegar que desde el externismo psicológico no puede conservarse el acceso privilegiado a los pensamientos pasados porque supone perder la capacidad de comparar la identidad/diferencia que existe entre los pensamientos pasados y presentes. Las respuestas que los partidarios del externismo ofrecieron ante este desafío fueron, como vimos, de muy distinto signo. Mientras que Ludlow y Tye admiten que el traslado a un entorno externo diferente conlleva la incapacidad de tener pensamientos que involucran contenidos adquiridos en el entorno pasado, Gibbons y Burge optan por mantener que el cambio de entorno no impide preservar los conceptos del entorno de origen. Por otra parte, muchos partidarios del compatibilismo defendieron que el acceso privilegiado a los propios contenidos mentales no incluye la capacidad de comparar entre contenidos pensados en distintos momentos del tiempo (Goldberg, Burge, Bernecker, Falvey y Owens). De ser así, aunque los casos de traslado socavaran dicha capacidad comparativa, ello no supondría, en último término, un atentado contra el acceso privilegiado *lato sensu*. Por nuestra parte, aunque dedicamos un espacio considerable al examen de estas propuestas y a la discusión de sus virtudes y defectos, desarrollamos el núcleo argumentativo de este apartado en torno a tres elementos, a saber, (i) la irrelevancia efectiva de las distinciones entre contenidos que establece el partidario del externismo (estrictamente) natural en los casos de traslado, (ii) algunos aspectos centrales de las propuestas de Moya y Grimaltós y (iii) la demanda de Ludlow de plantear casos de traslado realistas.

Con respecto al primer punto, hicimos notar que el cambio conceptual que, según esta clase de externismo, se produce en la víctima de los casos de traslado es un cambio con unos rasgos muy peculiares, al estar determinado por una diferencia en las propiedades subyacentes de los géneros naturales que resulta completamente desconocida para todos los hablantes del escenario en cuestión. Esto tiene como resultado que el cambio conceptual, en primer lugar, no es advertido *en el instante* en que se produce ni por el hablante, ni por ninguno de sus interlocutores posibles y tampoco es susceptible de ser detectado en ninguno de los *intercambios comunicativos* entre este agente y los demás hablantes, ya se sitúen estos en su entorno de origen o en el de acogida. En segundo lugar, se trata de una distinción que no sirve para aclarar ningún malentendido en la comunicación, ni conlleva diferencia alguna en el comportamiento del hablante que sufre el traslado. No obstante, la razón más concluyente para rechazar que se haya producido un cambio conceptual en el sujeto que sufre la clase de traslado que se discute en la mayor parte de la literatura al uso (que involucra dos entornos en los que la diferencia en los géneros naturales permanece oculta a todos los hablantes y no se incorpora a los usos socialmente asociados al término relevante) es que se trata de una variación que no sólo le resulta inadvertida al protagonista de los casos de traslado desde su perspectiva de primera persona, sino que también queda oculta a los interlocutores que observan su comportamiento lingüístico y no lingüístico. Con otras palabras, una diferencia en los conceptos que un hablante expresa en distintos momentos debería ser una diferencia dotada de un mínimo potencial explicativo, que tendría que ponerse de manifiesto en el comportamiento de este hablante o que habría de revelarse en los intercambios comunicativos que mantiene con sus interlocutores. Sin embargo, la variación en el repertorio conceptual que presuntamente se produce en los casos de traslado al uso no tiene ninguna de estas características, lo cual constituye un buen motivo para dudar de que dicho cambio haya tenido lugar.

En lo que concierne al punto (ii), pudimos extraer algunas sugerencias valiosas de los trabajos de Moya y Grimaltós sobre los casos de traslado, trabajos que en muchos aspectos se distancian de las propuestas compatibilistas con mayor éxito histórico. Así, suscribimos plenamente las reticencias de Moya ante las comprensiones causales “crudas” del externismo y destacamos una de sus observaciones, tan intuitiva como clarificadora, a saber, que la deferencia lingüística no es incondicional, lo cual supone que, por lo general, un sujeto considera que la comunidad lingüística en la que lleva a

cabo el aprendizaje de su lengua materna constituye la *comunidad de deferencia primaria*. De esta manera, cuando un hablante se traslada a un nuevo entorno, es decir, a un entorno en el que se habla un lenguaje distinto total o parcialmente, seguirá recurriendo por defecto a los usos lingüísticos de su comunidad de origen cuando un interlocutor de su comunidad de acogida le interrogue acerca de lo que él ha querido decir. Sin embargo, a pesar de que esta idea nos parece esencialmente correcta, nos mostramos escépticos con respecto a la utilidad que pudiera tener para afrontar los casos de traslado al uso, es decir, los casos que tienen lugar en el escenario de la Tierra Gemela descrito por Putnam. El motivo es que se trata de un escenario en el que no existe ninguna diferencia en los usos comunitarios asignados al término “agua” en un entorno y otro y donde parece difícil que un interlocutor de la Tierra Gemela interrogue alguna vez a la víctima de traslado sobre lo que él quiere decir cuando expresa sus pensamientos utilizando dicho término. En esta medida, acogerse primariamente a las convenciones lingüísticas de la comunidad de origen, como decimos, no sirve de ayuda para aclarar qué pensamiento ha querido expresar la víctima de traslado.

En lo que se refiere a las aportaciones de Grimaltós, este autor distingue, por un lado, los usos referenciales de los términos, en los que destaca la contribución del entorno natural y que son los usos predominantes en las aplicaciones demostrativas del término “agua” al líquido relevante, de los usos deferenciales, por otro, en los que prima la contribución del entorno social. De este modo, si la víctima de traslado hace un uso referencial del término “agua”, al designar mediante dicho término el líquido relevante de la Tierra Gemela, estará expresando un concepto distinto del concepto que expresaría al hacer un uso deferencial del término, acogiéndose a las convenciones socialmente asignadas al término en su comunidad de origen. No obstante, al igual que en el caso de Moya, dudamos del potencial de esta sugerencia en relación a los casos de traslado usuales, puesto que en esta clase de situaciones ambas comunidades de deferencia establecen unas convenciones idénticas para el uso del término “agua”, por lo que la contribución del entorno social en un caso y otro no determina diferencia alguna en cuanto al concepto expresado en los usos deferenciales del término. Ahora bien, Grimaltós nos proporciona una pista importante para afrontar los casos de traslado. Cuando una de estas víctimas es transportada a un entorno en el que se habla un lenguaje distinto del lenguaje que se habla en su comunidad de origen es frecuente que se produzca una *ruptura* entre el contenido de los pensamientos del hablante (lo que él pretende decir) y el contenido semánticamente expresado por los términos del lenguaje

de la comunidad de acogida, que será el contenido que figura en las atribuciones psicológicas que los hablantes de esa nueva comunidad realizarán para explicarse el comportamiento de la víctima del traslado. En relación con esto, también enumeramos y examinamos otros fenómenos lingüísticos (*lapsus*, falsos amigos) en los que se pone de manifiesto la existencia de una ruptura localizada entre el contenido pensado y el contenido semánticamente expresado por los hablantes. El hecho de calificar de “localizadas” a estas rupturas se deriva de nuestro convencimiento de que, por lo general, en las preferencias asertivas que llevan a cabo los hablantes ese contenido es uno y el mismo. La razón principal para asumir este principio es que si reconociéramos un carácter ubicuo a esas rupturas, ello nos llevaría tratar a los hablantes individuales como usuarios del lenguaje *à la* Humpty Dumpty, esto es, como agentes que constantemente rechazan la manera en que otros interpretan sus palabras alegando que, en realidad, eso no es lo que ellos querían decir (o lo que ellos estaban pensando) e impidiendo con ello el desempeño de los procesos comunicativos e interpretativos.

En tercer lugar, hicimos nuestra la demanda de Ludlow (presente también en los trabajos de Moya) de plantear casos de traslado realistas, que pueden afectar a usuarios del lenguaje efectivos en entornos actuales, sin necesidad de apelar a transportes inadvertidos de carácter inter-mundano. En este sentido, si lo que nos interesa es resolver puzzles sobre nuestras atribuciones ordinarias de pensamientos y aclarar nuestras intuiciones al respecto, parece que la propuesta de examinar casos de traslado con verdadera relevancia *para nosotros* es una estrategia que resultará verdaderamente útil en este sentido. De acuerdo con ello, cabe pensar, con Ludlow, que las situaciones efectivas que más se asemejan a los casos de traslado discutidos en la literatura filosófica serán aquellos en los que un hablante perteneciente a la comunidad lingüística CL_1 se traslada a una segunda comunidad lingüística, CL_2 , en la que se habla un lenguaje distinto pero parcialmente coincidente con el lenguaje hablado en CL_1 (una variante dialectal, por ejemplo). El motivo es que en esta situación se cumplen los requisitos básicos de los casos de traslado, a saber, en la comunidad de origen y en la comunidad de acogida se hablan distintos lenguajes pero parcialmente homófonos, de tal manera que al menos un término general “T” (en torno al cual se plantea el caso de traslado) tiene un significado diferente en la comunidad CL_1 y en la comunidad CL_2 . Asimismo, la víctima de traslado ha de desconocer, en parte, el significado que el término “T” tiene en su comunidad y, además, debe ignorar el hecho de que en la comunidad de acogida el término “T” tiene un significado distinto del que tiene en su

comunidad. A la previsible objeción de que un traslado de esta clase a otra comunidad lingüística no es un transporte *inadvertido* respondimos que el efecto de plantear un caso de traslado verdaderamente relevante es, precisamente, que nos lleva a diseñar una situación en la que un hablante cambia de comunidad lingüística a sabiendas, más bien que un secuestro repentino a un mundo cualitativamente idéntico, poblado por *Doppelgängern* de los hablantes del mundo actual. Sin embargo, lo importante para construir el caso de traslado se conserva, ya que nuestro hablante ignora la diferencia de significado comunitario del término “T” en un entorno y en otro. Lo que ocurrirá en un caso como este es, probablemente, que cuando nuestra víctima de traslado comunique un pensamiento en cuya expresión figura el término “T”, sus interlocutores de CL₂ le atribuirán un contenido correspondiente al expresado semánticamente por el término “T”, interpretado según los estándares de CL₂. Decimos “probablemente” porque también podría suceder que estos interlocutores conocieran la diferencia semántica que existe en ambas comunidades con respecto al término “T” y dudaran entre atribuirle el contenido semánticamente expresado según los estándares de CL₁ (que ellos conocen), o bien el expresado por los estándares de CL₂ o bien, simplemente, podrían preguntarle a nuestro hablante qué ha querido decir. Sin embargo, en caso de que los habitantes de CL₂ ignoraran también esa diferencia de significado, el contenido del pensamiento que le atribuirían a la víctima de traslado es, como decimos, idéntico al expresado semánticamente por la oración que este hablante ha utilizado para comunicar su pensamiento, según los estándares de CL₂. En tal caso, habría tenido lugar el fenómeno que Grimaltós diagnostica como una *ruptura* del contenido, es decir, que el contenido del pensamiento que nuestro hablante quiere expresar sería distinto del contenido semánticamente expresado por la oración que profiere para expresarlo (interpretada según los estándares de CL₂). Dicha ruptura se produce porque, como señala Moya, un hablante, por lo general, considerará a su comunidad lingüística de origen como su comunidad de deferencia primaria. Para comprobarlo, concebimos el supuesto de que este hablante es plenamente competente con respecto al uso del término “T” según los estándares de su comunidad y expresa un pensamiento mediante una oración en la que figura “T”. En este caso, si sus interlocutores le interpretaran haciendo uso de las convenciones lingüísticas propias de CL₂ (y existiera, por ello, una ruptura del contenido), nuestro protagonista podría negociar la interpretación que desea recibir. Como dijimos, es preferible hablar de “negociación” más que de rectificación porque la interpretación que él recibe no es exactamente incorrecta (es perfectamente correcta

para los habitantes de CL₂). No obstante, en la medida en que la atribución pretende capturar el contenido del pensamiento expresado por nuestro hablante, los habitantes de CL₂ deben aceptar que su interpretación de lo que la víctima del traslado quería decir no ha cumplido satisfactoriamente su objetivo y ha de concedérsele *el derecho* a proponer una interpretación distinta. Por nuestra parte, creemos que tal derecho existe en estas situaciones y otras similares (*lapsus*, falsos amigos) porque no estamos tratando con un agente que ha escogido caprichosamente una serie de palabras para decir algo muy distinto a lo que dichas palabras expresan semánticamente (uso del lenguaje *à la* Humpty Dumpty), sino que hablamos de alguien que tiene potestad para replicar “pensé que esas palabras expresaban para vosotros, en vuestro lenguaje, lo mismo que expresan para nosotros, en nuestro lenguaje”.

De manera similar, cuando un hablante no es plenamente competente en el significado de un término “T”, tal y como ocurre en los casos de traslado, y su uso del término es interpretado en su entorno de acogida CL₂ de una manera diferente a como sería interpretado por los hablantes de su comunidad de origen, no pensamos que haya ningún motivo para proceder generalmente de manera distinta. Si la interpretación que recibe se adecuara a cierto uso que la comunidad CL₂ da al término y que nuestro hablante (erróneamente) cree correcto en relación a las convenciones de su comunidad CL₁, la víctima podría no advertir en ese momento la diferencia existente entre lo que él ha querido decir y el contenido semánticamente expresado por sus palabras. Sin embargo, aunque dicha ruptura no fuera inmediatamente detectada sí es susceptible de serlo si un tercer hablante advierte este desfase, o bien si la víctima consultara un diccionario, o bien si entablara una conversación con algún experto relevante. Lo que nos interesa señalar es que, en cualquier caso, el hablante sabe lo que quiere decir, al igual que el paciente del experimento mental de Burge, porque tras ser consciente de la ruptura se diría “me he equivocado, aunque ellos, mis interlocutores de CL₂, creyeran que no ha sido así”. Lo importante es, en suma, que este sujeto tendría una serie de recursos a su alcance para advertir la ruptura de contenido.

Por otro lado, los casos de traslado realistas son situaciones bastante diferentes de aquella situación en la que se encuentra el protagonista de los casos de traslado al uso, desplegados en el primer experimento mental de la Tierra Gemela de Putnam, donde el término “T” en torno al cual se plantea el caso de traslado es un término de género natural que designa géneros esencialmente distintos en un entorno y otro y cuyo significado se individua en virtud de las propiedades subyacentes que poseen los

ejemplares de los géneros. En este caso, si admitiéramos que “T” posee un significado distinto en ambos entornos, en función de la diferencia en estas propiedades subyacentes, aun cuando los estándares comunitarios que regulan el uso del término fueran los mismos en las dos comunidades, nos encontraríamos con una situación en la cual el hablante usa el término “T” de manera perfectamente admisible para todos (tanto para los hablantes de CL₁ como para los hablantes de CL₂) y, sin embargo, se alega que, en realidad, los hablantes de CL₂ no han dado una interpretación adecuada de las palabras del sujeto que ha sufrido el traslado. Ahora bien, los medios para detectar esta presunta ruptura del contenido pensado y el contenido expresado semánticamente son muy distintos de aquellos que existen en los casos de traslado realistas. Así, sólo tras una serie de investigaciones científicas que podrían prolongarse durante siglos seríamos capaces de obtener el conocimiento de la diferencia en las propiedades subyacentes. Mientras tanto, ni este hablante, ni los interlocutores con los que él entra en contacto tanto en CL₁ como en CL₂, ni los diccionarios que puede consultar, etc., serían de ayuda para detectar esa ruptura de contenido. De admitir que existe un conflicto entre la interpretación que los hablantes de CL₂ realizan de las palabras del sujeto que ha sufrido el traslado y lo que éste en realidad quiere decir, se trataría de un conflicto muy peculiar, puesto que ni él ni nadie tendría noticia de que se ha producido hasta que se hubiera desarrollado la ciencia relevante.

La posibilidad de compatibilizar el externismo psicológico con el acceso privilegiado a los propios pensamientos pasa por descartar la viabilidad de la variante de externismo que se pone en juego en los casos de traslado usuales, variante que en un primer momento del trabajo describimos como “externismo natural” y cuyo alcance precisamos en el último capítulo del trabajo al caracterizar *la influencia natural directa*. En esta medida, pensamos que los factores externos relevantes para la fijación de la identidad de nuestros conceptos son únicamente las condiciones de aplicación socialmente determinadas para ellos y, especialmente, las establecidas por aquellos hablantes que son reconocidos socialmente como expertos. De ser esto cierto, no parece posible, como alegan los incompatibilistas, que yo esté queriendo expresar mediante el uso del término “T” cualquier concepto (*C*, *C'*, etc.) cuyas condiciones de aplicación (también determinadas socialmente) coincidan con las condiciones de aplicación del concepto *C* que yo sí conozco. La razón es que, por lo general, pretendo estar pensando en lo mismo que piensan los expertos en quienes confío yo y confían los demás hablantes de mi comunidad cuando hacemos uso del término “T” y no en lo que

pensaría ningún otro grupo de expertos de cualquier otro entorno, ya sea actual o contrafáctico.

El cuarto capítulo estuvo dedicado al examen del argumento incompatibilista de la consecuencia, un planteamiento relativamente complejo en el que se encuentran entrelazados otros problemas filosóficos como el escepticismo o la cuestión de qué hace que los argumentos lógicamente válidos sean, además de válidos, convincentes. Las distintas versiones del planteamiento de la consecuencia pretendían mostrar que, de mantener conjuntamente el externismo psicológico y el acceso *a priori* a los propios pensamientos, podríamos obtener conocimiento, igualmente *a priori*, de aquellos hechos externos de carácter empírico que determinan la identidad de nuestros contenidos mentales de acuerdo con el externismo. Pero, según dijimos a este respecto, dada la concepción de la aprioricidad puesta en juego en el argumento incompatibilista, resulta muy cuestionable tanto la traducción del privilegio epistémico de primera persona en conocimiento *a priori* de los propios pensamientos, como la tesis de que el externismo psicológico es una teoría planteada *a priori*. En relación a esto último, mostramos que el externismo es una tesis que cobra plausibilidad a partir de dos supuestos fundamentales, ninguno de los cuales está exento de material empírico, a saber, (i) el diseño de experimentos mentales que toman en consideración una serie de creencias empíricas sobre las prácticas científicas y lingüísticas efectivas, así como (ii) la atención a fenómenos empíricos relativos a nuestras prácticas de atribución de pensamientos y a los hábitos deferenciales presentes en una comunidad de hablantes. Por ello pensamos que las premisas que figuran en los distintos planteamientos del argumento de la consecuencia no son premisas establecidas puramente *a priori*, como se pretende, por lo que la conclusión extraída de ese argumento tampoco mantiene ese carácter. Asimismo, este hecho explica que cuando el argumento de la consecuencia se formula con propósitos anti-escépticos dicho razonamiento carece de poder convincente con respecto a las dudas que plantea el escéptico. De este modo, en la medida en que este hablante pretende que le convenzamos de que es racional escoger el primer miembro de la disyunción “existe el mundo externo o no existe el mundo pero todo me parece igual desde mi punto de vista”, no es eficaz plantear un argumento con este fin que emplee una premisa (el externismo) que él no aceptaría de partida, dada su clase de dudas inicial.

Finalmente, en el último capítulo de nuestro trabajo dedicamos una primera parte a distinguir aquellas tesis relativas a varios aspectos de nuestros procesos mentales

y estados psicológicos de carácter proposicional que se han atribuido a sí mismas el nombre de “externismo psicológico”. Aunque desde el principio dejamos claro que la clase de externismo cuya compatibilidad con el acceso privilegiado se discute es sólo el externismo relativo a la individuación de nuestros contenidos mentales, tuvimos ocasión de comprobar que en el debate sobre el compatibilismo algunos autores (compatibilistas e incompatibilistas) habían denominado “externismo psicológico” a otras teorías que no conciernen estrictamente a la cuestión de la identidad de los contenidos. En segundo lugar, nos hicimos cargo de las distintas variantes que usualmente se distinguen dentro del externismo con respecto a la individuación de nuestros contenidos mentales. A pesar de que en el primer capítulo fuimos capaces de extraer una comprensión básica de esta clase de externismo a partir del examen de los experimentos mentales de Putnam y Burge, creímos conveniente analizar el cuadro de las posibles combinaciones de tesis externistas en cuanto a la clase de factores externos cuya influencia reconocen. Nuestro análisis nos permitió poner en evidencia que la tipología usual de las variantes de externismo presenta ciertas tensiones internas, especialmente en el seno de una comprensión fuerte o amplia del externismo, que da cabida tanto a la influencia natural como a la influencia social en las tareas de individuación. El motivo es que en ocasiones los factores naturales, por un lado, y los factores sociales, por otro, empujan nuestras intuiciones en direcciones opuestas. Esta situación tiene lugar, primeramente, cuando el concepto expresado por el término “T” recibe una individuación más fina si atendemos a los factores naturales del entorno de los hablantes que si lo individualizamos de acuerdo con los usos socialmente asignados al término. En segundo lugar, también puede percibirse una tensión de esta clase cuando el concepto recibe una individuación más fina si atendemos a su significado público, comunitariamente establecido, antes que a los factores de tipo natural. Tal y como argumentamos, la existencia de estas dos situaciones revela que aquellas variedades de externismo que reconocen la influencia tanto de los factores naturales como de los factores sociales en la individuación de nuestros contenidos han de organizar su influencia y elegir cuál de ellos es el determinante en última instancia. A partir de estas conclusiones propusimos una nueva clasificación de las variedades de externismo que tuviera como objetivo evitar en la medida de lo posible las tensiones entre estas dos clases de elementos externos.

El segundo objetivo que nos propusimos en este último capítulo consistió en argumentar cuál es la variedad de externismo con respecto a la individuación de nuestros estados mentales que, con el mínimo atentado contra el acceso privilegiado,

realiza más aportaciones en cuanto a la explicación de nuestras prácticas cotidianas de atribución psicológica. En nuestra opinión, esta variedad de externismo tendría que subordinar la influencia de los factores naturales a la influencia de los elementos sociales en la individuación de nuestros contenidos, al mismo tiempo que toma en consideración los factores naturales de una manera mucho menos restrictiva que el externismo natural al uso. En este sentido, dedicamos cierto esfuerzo a explicar las insuficiencias de la teoría de los géneros naturales que subyace *tanto* al externismo estrictamente natural, esto es, el externismo que legitima la influencia individuadora de los factores naturales aun cuando ésta no queda incorporada de ninguna manera a los usos sociales, *como* a aquellas variantes de externismo – como la de Burge – que, aunque parecen otorgar una posición preeminente al uso convencional de los términos, siguen reservando un papel a la influencia de los factores naturales concebidos como propiedades subyacentes de los géneros. Con este fin, comenzamos presentando aquellos problemas que socavan la plausibilidad del escenario de la Tierra Gemela de Putnam y su dictamen sobre la referencia del término “agua”. Asimismo, nos hicimos cargo de las principales dificultades que aquejan a esta visión de los géneros naturales en el terreno de las especies biológicas, en particular, y en las prácticas taxonómicas actualmente vigentes en la ciencia, en general. Por último, cuestionamos la visión de Putnam del fenómeno de la deferencia lingüística, que es el elemento teórico que sustenta las atribuciones a hablantes no-expertos, ya transcurra su vida con anterioridad o con posterioridad a los presuntos descubrimientos de las propiedades subyacentes de los géneros. En todo caso, argumentamos que, aun cuando estos inconvenientes logran solucionarse o minimizarse, todavía sería necesario explicar en qué nos ayuda esta teoría de la individuación psicológica a la explicación de nuestros hábitos de atribución mental, puesto que, como hemos dicho, dicha teoría establece ciertas distinciones en los estados mentales de los hablantes que carecen de cualquier tipo de relevancia práctica, que no ayudan a resolver ningún puzzle o malentendido y que podrían pasar perfectamente desapercibidas para cualquier hablante en el caso de que los supuestos descubrimientos de propiedades subyacentes no llegaran a producirse nunca. Por este mismo motivo pensamos que las variedades de externismo que se sustentan total o parcialmente en esta teoría de los géneros naturales resultan especialmente agresivas con el acceso privilegiado de los hablantes a sus propios pensamientos. Así, en tanto que la identidad/diferencia de nuestros contenidos mentales depende de propiedades subyacentes de los géneros naturales que puede desconocer

tanto el sujeto que recibe la atribución como sus interlocutores, estaríamos reservando una influencia individuadora a factores que, de ser accesibles, sólo lo serían tras una cantidad ingente de trabajo empírico y cuya accesibilidad, además, no estaría garantizada en ningún momento, puesto que dicho trabajo es perfectamente falible. Si, por otro lado, esta teoría fuera asumida en su variante causal más cruda, obtendríamos que la identidad de los contenidos mentales de un sujeto puede sufrir multitud de cambios, en virtud de las interacciones que establece con factores naturales, sin que él o sus interlocutores adviertan nada en absoluto.

En lo que respecta a la teoría de la individuación que respaldamos, hemos rechazado aquellas consideraciones esencialistas presentes en la teoría de los géneros naturales que subyace al externismo natural en todas sus variantes, a favor de una actitud de tolerancia hacia los diversos puntos de vista en función de los cuales puede establecerse una clasificación de los géneros naturales. De este modo, los conceptos de género natural no tendrían el estatus especial que el externismo estrictamente natural les reserva, en la medida en que aquellos factores naturales que cuentan como criterios de individuación han de ser factores reconocidos por las convenciones sociales que regulan su uso. En cualquier caso, lo que cuestionamos es la relevancia de aquellas propiedades que *todos* los hablantes podrían desconocer en la individuación de nuestros contenidos mentales, propiedades que, de acuerdo con lo dicho, determinarían nuestros pensamientos independientemente de nuestro punto de vista, de nuestros intereses y de los intereses de todos nuestros interlocutores. Rechazamos esta visión, en suma, porque queremos abogar por una teoría de la individuación que realmente nos sirva para determinar y explicar qué estamos pensando y qué están pensando los demás.

El enfoque externista al que nos adherimos considera que la atribución a un hablante de pensamientos que involucran conceptos cuyas condiciones de aplicación él no domina (ni ignora) por completo es perfectamente legítima, en cuanto que dicho hablante se encuentra inserto en una comunidad de hablantes con los que pretende entenderse. Esta presuposición básica se concreta en relaciones de deferencia hacia uno o varios grupos de expertos, en el reconocimiento de la autoridad de sus correcciones y en el posible acceso a expedientes e instituciones con el fin de averiguar cómo *debo* decir aquello que *quiero decir* para que otros me entiendan. Pensamos, pues, que para mantener el acceso privilegiado a mis propios pensamientos y conservar mi potestad sobre lo que quiero decir, debo reconocermelo como usuario de un lenguaje socialmente regulado y, además, ha de mantenerse abierta la posibilidad de consultar a personas y

acceder a recursos que expliciten esas regulaciones. En definitiva, desde este punto de vista uso mis palabras para decir algo que quiero que lo demás comprendan y, en consecuencia, mis intenciones comunicativas han de someterse al arbitrio de significados públicos que, en parte, puedo desconocer, pero cuyo conocimiento me resulta potencialmente accesible en tanto que miembro de un cuerpo social.

Bibliografía

- ACERO, J. J. (1995): "Teorías del Contenido Mental". En Broncano (1995), pp. 362-393.
- ALMOG, J., J. PERRY, H. K. WETTSTEIN Y D. KAPLAN (eds.) (1989): *Themes from Kaplan*. Nueva York: Oxford University Press.
- ALLN, S. (ed.) (1989): *Possible Worlds in Humanities, Arts and Sciences*. Berlín: W. de Gruyter.
- ARDÉVOL PIERA, E., G. MUNILLA CABRILLANA Y J. CERVELLÓ AUTUORI (2003): *Antropología de la Religión: una Aproximación Interdisciplinar a las Religiones Antiguas y Contemporáneas*. Barcelona: Universitat Oberta de Catalunya.
- ATLAS, J. D. (2007): "Why Water Isn't H₂O, Much Less Necessarily H₂O?". En Davidson (2007), pp. 495-496.
- BAIRD, D., E. SCERRI Y L. MCINTYRE (eds.) (2006): *Philosophy of Chemistry: Synthesis of a New Discipline*. Dordrecht: Springer.
- BAKER, L. R. (2007): "Social Externalism and First Person Authority". *Erkenntnis*, 67, pp. 287-300.
- BEATTIE, J. (1993): "Is Putnam Inconsistent and Parochial?". *Australasian Journal of Philosophy*, 71, pp. 316-324.
- BEEBEE, H. (2001): "Transfer of Warrant, Begging the Question and Semantic Externalism". *The Philosophical Quarterly*, 51, pp. 356-374.
- BEEBEE, H. Y N. SAABARTON-LEARY (2010a): "On the Abuse on the Necessary *a posteriori*". En Beebee/Saabarton-Leary (2010b), pp. 159-178.
- BEEBEE, H. Y N. SAABARTON-LEARY (eds.) (2010b): *The Semantics and Metaphysics of Natural Kinds*. Nueva York/Londres: Routledge.
- BERNECKER, S. (1996): "Externalism and the Attitudinal Component of Self-Knowledge". *Nous*, 30, pp. 262-275.
- BERNECKER, S. (1997): "Self-Knowledge and Closure". En Ludlow/Martin (1998), pp. 333-349.
- BERNECKER, S. (2004): "Memory and Externalism". *Philosophy and Phenomenological Research*, 60, pp. 605-632.
- BHUSHAN, N. (2006): "Are Chemical Kinds Natural Kinds?". En Baird/Scerri/McIntyre (2006), pp. 327-336.
- BILGRAMI, A. (1992a): *Belief and Meaning: the Unity and Locality of Mental*. Cambridge: Blackwell.
- BILGRAMI, A. (1992b): "Can Externalism be Reconciled with Self-Knowledge?". En Pessin/Goldberg (1996), pp. 362-393.
- BILGRAMI, A. (2003): "A Trilemma for Redeployment". *Philosophical Issues*, 13, pp. 22-30.
- BIRD, A. (2010): "A *posteriori* Knowledge of Natural Kind Essences: A Defence". *Philosophical Topics*, 35, pp. 293-312.
- BOGHOSSIAN, P. (1989): "Content and Self-Knowledge". En Ludlow/Martin (1998), pp. 149-173.
- BOGHOSSIAN, P. (1992): "Externalism and Inference". *Philosophical Issues*, 2, pp. 11-28.
- BOGHOSSIAN, P. (1994): "The Transparency of Mental Content". *Philosophical Perspectives*, 8, pp. 33-50.
- BOGHOSSIAN, P. (1998): "What the Externalist Can Know '*a priori*'?". *Philosophical Issues*, 9, pp. 197-211.
- BOGHOSSIAN, P. y C. PEACOCKE (eds.) (2000): *New Essays on the a priori*. Oxford: Clarendon Press.

- BOYD, R. (1999a): "Homeostasis, Species, and Higher Taxa". En Wilson (1999a), pp. 141-186.
- BOYD, R. (1999b): "Kinds, Complexity and Multiple Realization". *Philosophical Studies*, 95, pp. 67-98.
- BREWER, B. (2000): "Externalism and *a priori* Knowledge of Empirical Facts". En Boghossian/Peacocke (2000), pp. 384-414.
- BRIGANDT, I. (2004): "Biological Kinds and the Causal Theory of Reference". En Marek/Reicher (2004), pp. 58-60.
- BRONCANO, F. (1995): *La Mente Humana*. Madrid: Trotta.
- BROWN, J. (1995): "The Incompatibility of Anti-individualism and Privileged Access". En Ludlow/Martin (1998), pp. 185-193.
- BROWN, J. (1999): "Boghossian on Externalism and Privileged Access". *Analysis*, 59, pp. 52-59.
- BROWN, J. (2004): *Anti-individualism and Knowledge*. Cambridge, Mass.: MIT Press.
- BROWN, J. (2009): "Semantic Externalism and Self-Knowledge". En McLaughlin/Beckermann/Walter (2009), pp. 767-780.
- BRUECKNER, A. (1986): "Brains in a Vat". *The Journal of Philosophy*, 83, pp. 148-167.
- BRUECKNER, A. (1992): "What an Anti-individualist Knows *a priori*". *Analysis*, 59, pp. 52-59.
- BRUECKNER, A. (1997a): "Externalism and Memory". En Ludlow/Martin (1998), pp. 319-331.
- BRUECKNER, A. (1997b): "Is Scepticism about Self-Knowledge Incoherent?". *Analysis*, 57, pp. 287-290.
- BRUECKNER, A. (2003): "The Coherence of Scepticism about Self-Knowledge". *Analysis*, 63, pp. 41-48.
- BRUECKNER, A. (2004): "McKinsey Redux?". En Schantz, pp. 377-388.
- BRUECKNER, A. (2007): "Externalism and Privileged Access are Consistent". En McLaughlin/Cohen (2007), pp. 33-52.
- BURGE, T. (1979): "Individualism and the Mental". En Burge (2007), pp. 100-180.
- BURGE, T. (1982): "Other Bodies". En Burge (2007), pp. 82-99.
- BURGE, T. (1986a): "Individualism and Psychology". *Philosophical Review*, 95, pp. 3-45.
- BURGE, T. (1986b): "Cartesian Error and the Objectivity of Perception". En Burge (2007), pp. 192-207.
- BURGE, T. (1986c): "Intellectual Norms and Foundations of Mind". En Burge (2007), pp. 254-274.
- BURGE, T. (1988a): "Individualism and Self-Knowledge". *Journal of Philosophy*, 85, pp. 649-663.
- BURGE, T. (1988b): "Authoritative Self-knowledge and Perceptual Individualism". En Burge (2007), pp. 208-220.
- BURGE, T. (1989a): "Wherein is Language Social". En Burge (2007), pp. 275-290.
- BURGE, T. (1989b): "Individuation and Causation in Psychology". En Burge 2007, pp. 316-333.
- BURGE, T. (1995): "Intentional Properties and Causation". En Burge (2007), pp. 334-343.
- BURGE, T. (1996): "Our Entitlement to Self-Knowledge". En Ludlow/Martin (1998), pp. 239-263.
- BURGE, T. (1998): "Memory and Self-Knowledge". En Ludlow/Martin (1998), pp. 351-370.

- BURGE, T. (2003a): "Reply to Carlos Moya: 'Externalism, Inclusion, and Knowledge of Content'". En Frápolli/Romero (2003), pp. 277-282.
- BURGE, T. (2003b): "Reply to Martin Davies: 'Externalism, Self-Knowledge and Transmission of Warrant'". En Frápolli/Romero (2003), pp. 105-130.
- BURGE, T. (2006): "Postscript to 'Individualism and the Mental'". En Burge (2007), pp. 151-181.
- BURGE, T. (2007): *Foundations of Mind*. Nueva York: Oxford University Press.
- CAMPBELL, J., M.O. ROURKE Y D. SHIER (eds.) (2002): *Meaning and Truth*. Nueva York: Seven Bridges Press.
- CARNAP, R. (1955): "Meaning and Synonymy in Natural Languages". En Carnap (1956), pp. 233-246.
- CARNAP, R. (1956): *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.
- CARRUTHERS, P. (1989): *Introducing Persons: Theories and Arguments in the Philosophy of Mind*. Londres: Routledge.
- CASTAÑEDA, H-N. (1989): "Semantic Holism without Semantic Socialism: Twin Earths, Thinking, Language, Bodies and the World". En French/Uehling/Wettstein (1989), pp. 101-126.
- CLARIDGE, M.F., H.A. DAWAH Y M.R. WILSON (eds.) (1997): *Species: the Units of Biodiversity*. Nueva York: Springer-Verlag.
- CLARK, P. Y B. HALE (eds.) (1994): *Reading Putnam*. Cambridge: Blackwell.
- CORBÍ, J. E. (1998): "A Challenge to Boghossian's Incompatibilist Argument". *Philosophical Issues*, 9, pp. 231-242.
- CRANE, T. (2001): *Elements of Mind*. Nueva York: Oxford University Press.
- CRIMMINS, M Y J. PERRY (1989): "The Prince and the Phone Booth: Reporting Puzzling Beliefs". En Davidson (2007), pp. 356-379.
- CHALMERS, D. (1996): *The Conscious Mind: in Search of a Fundamental Theory*. Nueva York: Oxford University Press.
- CHISHOLM, R. (1974): "On the Nature of Acquaintance: A Discussion of Russell's Theory of Knowledge". En Irvine (1999), pp. 211-220.
- CHOMSKY, H. (1995): "Language and Nature". *Mind*, 104, pp. 1-61.
- DAVIDSON, D. (1973): "Radical Interpretation". En Davidson (1984), pp. 125-139.
- DAVIDSON, D. (1965): "Theories of Meaning and Learnable Languages". En Davidson (1984), pp. 3-15.
- DAVIDSON, D. (1967): "Truth and Meaning". En Davidson (1984), pp. 17-36.
- DAVIDSON, D. (1974): "On the Very Idea of a Conceptual Scheme". En Davidson (1984), pp. 183-198.
- DAVIDSON, D. (1983): "A Coherence Theory of Truth and Knowledge". En Davidson (2004), pp. 137-153.
- DAVIDSON, D. (1984): "First Person Authority". En Davidson (2001), pp. 3-14. (Trad. cast.: "La Autoridad de Primera Persona", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 25-40).
- DAVIDSON, D. (1984): *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
- DAVIDSON, D. (1986): "A Nice Derangement of Epitaphs". En Davidson (2006), pp. 89-108.
- DAVIDSON, D. (1987): "Knowing One's Own Mind". En Davidson (2001), pp. 15-38. (Trad. cast.: "Conocer Nuestra Propia Mente", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 41-71).

- DAVIDSON, D. (1988a): "The Myth of the Subjective". En Davidson (2001), pp. 39-52. (Trad. cast.: "El Mito de lo Subjetivo", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 72-90).
- DAVIDSON, D. (1988b): "Reply to Burge". *The Journal of Philosophy*, 85, pp. 664-665.
- DAVIDSON, D. (1990): "Epistemology Externalized". En Davidson (2001), pp. 205-220. (Trad. cast.: "La Externalización de la Epistemología", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 265-279).
- DAVIDSON, D. (1991): "Three Varieties of Knowledge". En Davidson (2001), pp. 205-220. (Trad. cast.: "Tres Variedades de Conocimiento", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 193-204).
- DAVIDSON, D. (1992): *Mente, Mundo y Acción: Claves para una Interpretación*. Barcelona: Paidós.
- DAVIDSON, D. (1997): "Indeterminism and Antirealism". En Davidson (2001), pp. 69-84. (Trad. cast.: "El Indeterminismo y el Antirrealismo", en Davidson, *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003, pp. 110-129).
- DAVIDSON, D. (2001): *Subjective, Intersubjective, Objective*. Oxford: Oxford University. (Trad. cast.: *Subjetivo, Intersubjetivo, Objetivo*, Madrid, Cátedra, 2003).
- DAVIDSON, D. (2006): *Truth, Language and History*. Oxford: Oxford University Press.
- DAVIDSON, M. (ed.) (2007): *On Sense and Direct Reference. Readings in the Philosophy of Language*. Boston: McGraw-Hill.
- DAVIES, M. (1998): "Externalism, Architecturalism, and Epistemic Warrant". En Wright/Smith/MacDonald (2000), pp. 321-361.
- DAVIES, M. (2000): "Externalism and Armchair Knowledge". En Boghossian/Peacocke (2000), pp. 384-414.
- DAVIES, M. (2003a): "Externalism, Self-knowledge and Transmission of Warrant". En Frápolli/Romero (2003), pp. 105-131.
- DAVIES, M. (2003b): "The Problem of Armchair Knowledge". En Nuccetelli (2003b), pp. 23-56.
- DAVIS, S. (1983): *Causal Theories of Mind: Action, Knowledge, Memory, Perception and Reference*. Berlín: Walter De Gruyter.
- DESCARTES, R. (1641): *Méditations Métaphysiques*. París: Presses Universitaires de France. (Trad. cast: *Discurso del método; Meditaciones metafísicas*, Madrid, Espasa-Calpe, 1975).
- DEVITT, M. (1981): *Designation*. Nueva York: Columbia University Press.
- DEVITT, M. (1989): "Against Direct Reference". En Davidson (2007), pp. 462-494.
- DEVITT, M. (2011): "Deference and The Use Theory". *Protosociology: An International Journal of Interdisciplinary Research*, 27, pp. 196-211.
- DEVITT, M. Y K. STERELNY (1999): *Language and Reality. An Introduction to the Philosophy of Language*. Cambridge, Mass.: MIT Press, 2ª ed., rev. y ampliada; 1ª ed., 1987.
- DONNELLAN, K. (1966): "Reference and Definite Descriptions". En Schwartz (1990), pp. 42-65.
- DONNELLAN, K. (1993): "There is a Word for that Kind of Thing: An Investigation of Two Thought Experiments". *Philosophical Perspectives*, 7, pp. 155-168.
- DONNELLAN, K. (2003): "Burge Thought Experiments". En Hahn/Ramberg (2003), pp. 59-75.
- DOWTY, D. R. (1979): *Word, Meaning and Montague Grammar*. Reidel: Dordrecht.
- DRETSKE, F. (1970): "Epistemic Operators". *Journal of Philosophy*, 67, pp. 1007-1023.
- DRETSKE, F. (2003): "Externalism and Self-Knowledge". En Nuccetelli (2003b), pp. 131-142.

- DRETSKE, F. (2005): "The Case against Closure". En Steup/Sosa (2005), pp. 13-25.
- DUMMETT, M. (1993): *Frege: Philosophy of Language*. Cambridge: Harvard University Press.
- DUPRÉ, J. (1981): "Natural Kinds and Biological Taxa". *Philosophical Review*, 90, pp. 66-90.
- DUPRÉ, J. (1993): *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge Mass.: Cambridge University Press.
- DUPRÉ, J. (1996): "Promiscuous Realism: Reply to Wilson". *The British Journal for the Philosophy of Science*, 47, pp. 441-444.
- DUPRÉ, J. (2002): "Is 'Natural Kind' a Natural Kind Term?". *The Monist*, 85, pp. 29-49.
- DUPRÉ, J. (2004): "Review of Joseph LaPorte, *Natural Kinds and Conceptual Change*". *Notre Dame Philosophical Reviews*, 6. URL=<<http://ndpr.icaap.org/content/archives/2004/6/duprelaporte.html>>.
- DUPRÉ, J. (2006): *Humans and Other Animals*. Oxford: Clarendon.
- EBBS, G. (1996): "Can We Take Our Words at Face Value?". *Philosophy and Phenomenological Research*, 56, pp. 499-530.
- EBBS, G. (2001): "Is Scepticism about Self-Knowledge Coherent?". *Philosophical Studies*, 105, pp. 43-58.
- EBBS, G. (2003): "A Puzzle about Doubt". En Nuccetelli (2003b), pp. 131-142.
- EBBS, G. (2005): "Why Skepticism about Self-Knowledge is Self-Undermining". *Analysis*, 65, pp. 237-244.
- EDWARDS, P. (ed.) (1972): *The Encyclopaedia of Philosophy*, vol. 6. Nueva York: MacMillan.
- ELLIS, B. (2001): *Scientific Essentialism*. Cambridge: Cambridge University Press.
- ELLIS, B. (2002): *Philosophy of Nature: A Guide to the New Essentialism*. Chesham: Acumen.
- ERESHEFSKY, M. (2010a): "Microbiology and the Species Problem". *Biology and Philosophy*, 25, pp. 553-568.
- ERESHEFSKY, M. (2010b): "What's Wrong with the New Biological Essentialism". *Philosophy of Science*, 77, pp. 674-685.
- ESFELD, M. (2002): "Externalism about Content: Its Social and Physical Roots". *Filosoficky Casopis*, 50, pp. 387-400.
- EVANS, G. (1979): "Reference and Contingency". En Evans (1985), pp. 178-213.
- EVANS, G. (1982): *The Varieties of Reference*. Oxford: Clarendon Press.
- EVANS, G. (1985): *Collected Papers*. Oxford: Clarendon Press.
- FALVEY, K. (2003): "Memory and Knowledge of Content". En Nuccetelli (2003b), pp. 219-240.
- FALVEY, K. y OWENS, J. (1994): "Externalism, Self-knowledge, and Skepticism". *Philosophical Review*, 103, pp. 107-37.
- FARKAS, K. (2003): "What is Externalism?". *Philosophical Studies*, 112, pp. 187-208.
- FARKAS, K. (2008): *The Subject's Point of View*. Nueva York: Oxford University Press.
- FERNÁNDEZ MORENO, L. (2000): "¿Qué es un Término de Género Natural?". *Teorema*, XIX, pp. 45-58.
- FERNÁNDEZ MORENO, L. (2006): *La Referencia de los Nombres Propios*. Madrid: Trotta.
- FERNÁNDEZ MORENO, L. (ed.) (2009): *Language, Nature and Science: New Perspectives*. Madrid: Plaza y Valdés.
- FERNÁNDEZ MORENO, L. (2012): "Natural Kind Terms, Rigidity and the Path towards Necessity". *Acta Philosophica*, 21, pp. 109-122.

- FODOR, J. A. (1988): *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Mass.: MIT Press.
- FRÁPOLLI, M. J. y E. ROMERO (eds.) (2003): *Meaning, Basic Self-knowledge, and Mind: Essays on Tyler Burge*. Stanford: CSLI Publications.
- FREGE, G. (1892): "Über Sinn und Bedeutung". En G. Frege *Funktion, Begriff, Bedeutung*, Gotinga: Vandenhoeck & Ruprecht, 1979, pp. 40-65. (Trad. cast.: en L.M. Valdés (ed.), *La Búsqueda del Significado*, Madrid, Tecnos, 2005, pp. 29-49).
- FREGE, G. (1984): "Thoughts". En Salmon/Soames (1988), pp. 33-55.
- FRENCH, P., T. UEHLING, y H. WETTSTEIN (eds.) (1979): *Midwest Studies in Philosophy XIV. Contemporary Perspectives in the Philosophy of Language II*. Minneapolis: University of Minnesota Press.
- GALLAGHER, S. (2008): *The Phenomenological Mind: an Introduction to Philosophy of Mind and Cognitive Science*. Nueva York: Routledge.
- GARCÍA SUÁREZ, A. (2000): "On Wright's Diagnosis of McKinsey's Argument". *Philosophical Issues*, 10, pp. 164-171.
- GARCÍA SUÁREZ, A. (2011): *Modos de Significar: una Introducción Temática a la Filosofía del Lenguaje*. Madrid: Tecnos, 2ª ed., rev. y ampliada; 1ª ed., 1997.
- GIBBONS, J. (1996): "Externalism and Knowledge of Content". *The Philosophical Review*, 105, pp. 287-310.
- GOLDBERG, S. (1997): "Self-Knowledge, Self-Ascription, and the Memory Argument", *Analysis*, 57, pp. 211-219.
- GOLDBERG, S. (1999): "The Relevance of Discriminatory Knowledge of Content", *Pacific Philosophical Quarterly*, 80, pp. 136-156.
- GOLDBERG, S. (2003a): "What do You Know When You Know your Own Thoughts?". En Nuccetelli (2003b), pp. 241-256.
- GOLDBERG, S. (2003b): "On our Alleged *a priori* Knowledge that Water Exists". *Analysis*, 63, pp. 38-41.
- GOLDMAN, A. (1976a): "What is Justified Belief?". En Sosa/Kim (2000), pp. 340-353.
- GOLDMAN, A. (1976b): "Discrimination and Perceptual Knowledge". En Davis (1983), pp. 73-113.
- GRACIA, D. Y J. LÁZARO (2009): *Introducción a la Medicina: Historia, Teoría, Terminología y Documentación*. Madrid: Hariadna.
- GRICE, H. P. (1957): "Meaning". *The Philosophical Review*, 66, pp. 377-388. (Trad. cast.: en L.M. Valdés (ed.), *La Búsqueda del Significado*, Madrid, Tecnos, 2005, pp. 481-490).
- GRIMALTÓS, T. (2003a): "Terms and Content". En Frápolli (2003), pp. 35-52.
- GRIMALTÓS, T. (2003b): "Cerebros y Escarabajos. Sobre el Argumento Anti-escéptico de Putnam". *Teorema*, 20, pp. 21-30.
- GUTTENPLAN, S. (2000): *Mind's Landscape: an Introduction to the Philosophy of Mind*. Oxford: Blackwell.
- HACKING, I. (2007): "Putnam's Theory of Natural Kinds and their Names is not the same as Kripke's". *Principia*, 11, pp. 1- 24.
- HAHN, M. Y B. RAMBERG (eds.) (2003): *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge, Mass.: MIT Press.
- HALLETT, G. L. (1991): *Essentialism: a Wittgensteinian Critique*. Albany: State University Press.
- HAMPTON, J. A. (1998). "Folk Biology and External Definitions". *Behavioral and Brain Sciences*, 21, pp. 573-574
- HANFLING, O. (2000): *Philosophy and Ordinary Language: the Bent and Genius of our Tongue*. Nueva York/Londres: Routledge.

- HEAL, J. (1988): "Externalism and Memory." *Proceedings of the Aristotelian Society*, 88, pp. 183–209.
- HEIL, J. (1988): "Privileged Access". *Mind*, 47, pp. 238-51.
- HEIL, J. (2004): *Philosophy of Mind: a Guide and Anthology*. Oxford: Oxford University Press.
- IRVINE, A. D. (ed.) (1999): *Bertrand Russell: Critical Assessments; Logic, Knowledge and the World*, vol. III. Nueva York: Routledge.
- JACKMAN, H. (2000): "Deference and Self-Knowledge". *Southwest Philosophy Review*, 16, pp. 171-180.
- JACKMAN, H. (2003): "Expression, Thought and Language". *Philosophia*, 31, pp. 33-54.
- JACKSON, F. Y P. PETTIT (1990): "In Defence of Folk Psychology". *Philosophical Studies*, 59, pp. 31-54.
- JAEGWON, K. (1998): *Philosophy of Mind*. Colorado: Westview Press.
- JAUME, A., M. LIZ, D. PÉREZ CHICO, M. PONTE Y M. VÁZQUEZ (2010): *Proceedings of the Sixth Conference of the Spanish Society for Analytic Philosophy*. Santa Cruz de Tenerife.
- KAPLAN, D. (1989a): "Demonstratives". En Almog/Perry/Wettstein/Kaplan (1989), pp. 481-564.
- KAPLAN, D. (1989b): "Afterthoughts". En Almog/Perry/Wettstein/Kaplan (1989), pp. 565-614.
- KORMAN, D. Z. (2006): "What an Externalist Would Say about Dry Earth". *Journal of Philosophy*, pp. 503-520.
- KRAAY, K. J. (2002): "Externalism, Memory, and Self-Knowledge". *Erkenntnis*, 56, pp. 297-317.
- KRIPKE, S. (1977): "Speaker's Reference and Semantic Reference". *Midwest Studies in Philosophy*, 2, pp. 255-276.
- KRIPKE, S. (1980): *Naming and Necessity*. Oxford: Blackwell.
- KRIPKE, S. (2011a): "Vacuous Names and Fictional Entities". En Kripke (2011b), pp. 254-291.
- KRIPKE, S. (2011b): *Philosophical Troubles. Collected Papers, Vol. 1*. Nueva York: Oxford University Press.
- KUHN, TH. S. (1989): "Possible Worlds in History of Science". En Allén (1989), pp. 9-32.
- LAPORTE, J. (2004): *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.
- LAPORTE, J. (2010): "Theoretical Identity Statements, Their Truth, and Their Discovery". En Beebe/Sabbarton-Leary (2010b), pp. 104-124.
- LEPORE, E. Y K. LUDWIG (2005): *Donald Davidson: Meaning, Truth, Language and Reality*. Oxford: Clarendon Press.
- LIU, J. (2002): "Physical Externalism and Social Externalism: Are They Really Compatible?". *Journal of Philosophical Research*, 27, pp. 381- 404.
- LOAR, B. (1985): "Social Content and Psychological Content". En Pessin/Goldberg (1996), pp. 180-190.
- LOCKE, J. (1690): *An Essay Concerning Human Understanding*. Chicago: University of Chicago Press. (Trad. cast.: en J. Locke, *Ensayo sobre el Entendimiento Humano*, Santa Fe de Bogotá, Fondo de cultura económica, 2008).
- LUDLOW, P. (1995a): "Externalism, Self-Knowledge, and the Prevalence of Slow Switching". En Ludlow/Martin (1998), pp. 225-230.
- LUDLOW, P. (1995b): "Social Externalism, Self-Knowledge, and Memory". En Ludlow/Martin (1998), pp. 307-310.

- LUDLOW, P. (1996): "Social Externalism and Memory: a Problem?". En Ludlow/Martin (1998), pp. 311-317.
- LUDLOW, P. (1997): "On the Relevance of Slow Switching". En Ludlow/Martin (1998), pp. 235-237.
- LUDLOW, P. (2004): "What Was I Thinking? Social Externalism, Self-Knowledge and Shifting Memory Targets". En Schantz (2004), pp. 419-425.
- LUDLOW, P. Y N. MARTIN (eds.) (1998): *Externalism and Self-Knowledge*. Stanford: CSLI Publications.
- LYCAN, G. W. (2006): "The Meaning of Water: An Unsolved Problem". *Philosophical Issues*, 16, pp. 184-199.
- MARCONI, D. (1998): *Lexical Competence*. Cambridge, Mass.: MIT Press. (Trad. cast.: *La Competencia Léxica*, Madrid, Visor, 2000).
- MAREK, J. C. Y M. E. REICHER (eds.) (2004): *Experience and Analysis: Papers of the 27th International Wittgenstein Symposium*. Austrian Ludwig Wittgenstein Society: Kirchberg am Wechsel.
- MARTÍ, G. (1995): "The Essence of Genuine Reference". *Journal of Philosophical Logic*, 24, pp. 275-289.
- MARTÍ, G. (2003): "The Question of Rigidity in New Theories of Reference". *Nous*, 37, pp. 161-179.
- MARTÍ, G. Y J. MARTÍNEZ-FERNÁNDEZ (2010): "General Terms as Designators: A Defence of the View". En H. Beebe/Saabarton-Leary (2010b), pp. 46-63.
- MARTINICH, A. (ed.) (1996): *The Philosophy of Language*. Oxford: Oxford University Press.
- MAYDEN, R. L. (1997): "A Hierarchy of Species Concepts: the Denouement in the Saga of Species Problem". En Claridge/Dawah/Wilson (1997), pp. 381- 424.
- MCDOWELL, J. (1984): "De Re Senses". *Philosophical Quarterly*, 34, pp. 283-294.
- MCDOWELL, J. (1992): "Putnam on Mind and Meaning". En Pessin/Goldberg (1996), pp. 305-318.
- MCGINN, C. (1977): "Charity, Interpretation and Belief". *The Journal of Philosophy*, 74, pp. 521-535.
- MCGINN, C. (1997): *The Character of Mind: An Introduction to the Philosophy of Mind*. Nueva York: Oxford University Press.
- MCKAY, T. (1981): "On Proper Names in Belief Ascriptions". En Davidson (2007), pp. 266-280.
- MCKINSEY, M. (1991): "Anti-individualism and Privileged Access". *Analysis*, 51, pp. 9-16.
- MCKINSEY, M. (2002a): "Forms of Externalism and Privileged Access". *Philosophical Perspectives*, 16, pp. 199-224.
- MCKINSEY, M. (2002b): "The Semantic Basis of Externalism". En Campbell/Rourke/Shier (2002), pp. 34-45.
- MCKINSEY, M. (2003): "Transmission of Warrant and Closure of Apriority". En Nuccetelli (2003b), pp. 97-116.
- MCKINSEY, M. (2007): "Externalism and Privileged Access are Inconsistent". En McLaughlin/Cohen (2007), pp. 53-65.
- MCLAUGHLIN, B., A. BECKERMANN Y S. WALTE (eds.) (2009): *The Oxford Handbook of Philosophy of Mind*. Nueva York: Oxford University Press.
- MCLAUGHLIN, B.P. (2003): "McKinsey's Challenge, Warrant Transmission, and Skepticism". En Nuccetelli (2003b), pp. 79-96.
- MCLAUGHLIN, B.P. y J. COHEN (eds.) (2007): *Contemporary Debates in Philosophy of Mind*. Malden: Blackwell.

- MCLAUGHLIN, B.P. y M. TYE (1998a): "Is Content-Externalism Compatible with Privileged Access?". *Philosophical Review*, 107, pp. 349-380.
- MCLAUGHLIN, B.P. Y M. TYE (1998b): "The Brown-McKinsey Charge of Inconsistency". En Ludlow/Martin (1998), pp. 207-214.
- MCLAUGHLIN, B.P., A. BECKERMANN y S. WALTER (eds.) (2009): *The Oxford Handbook of Philosophy of Mind*. Oxford: Oxford University Press.
- MELLOR, D. H. (1977): "Natural Kinds". En Pessin/Goldberg (1996), pp. 69-80.
- MENARY, R. (2010): *The Extended Mind*. Cambridge, Mass.: MIT Press.
- MONGE-NÁGERA, J., P. GÓMEZ FIGUEROA Y M. RIVAS ROSSI (2002): *Biología General*. San José: EUNED.
- MOORE, G. E. (1939): "Proof of an External World". En Moore (1970), pp. 127-150.
- MOORE, G. E. (1970): *Philosophical Papers*. Londres: George Allen and Unwin.
- MOYA, C. (1998): "Boghossian's *Reductio* of Compatibilism". *Philosophical Issues*, 9, pp. 243-251.
- MOYA, C. (2003): "Externalism, Inclusion, and Knowledge of Content". En Frápolli/Romero (2003), pp. 165-186.
- MOYA, C. (2004): *Filosofía de la Mente*. Valencia: Universitat de València.
- NAGEL, T. (1974): "What is Like to Be a Bat?". En Heil (2004), pp. 528-538.
- NEEDHAM, P. (2000): "What is Water?". *Analysis*, 60, pp. 13-21.
- NUCCETELLI, S. (1999): "What Anti-individualists Cannot Know *a priori*". *Analysis*, 59, pp. 48-51.
- NUCCETELLI, S. (2003a): "Knowing that One Knows What One is Talking About". En Nuccetelli (2003b), pp. 169-184.
- NUCCETELLI, S. (ed.) (2003b): *New Essays on Semantic Externalism and Self-knowledge*. Cambridge, Mass.: MIT Press.
- OKASHA, S. (2002): "Darwinian Metaphysics: Species and the Question of Essentialism". *Synthese*, 131, pp. 191-213.
- OWENS, J. (1987): "In Defence of a Different Doppelgänger". *The Philosophical Review*, 96, pp. 521-554.
- OWENS, J. (2003): "Externalism, Davidson and Knowledge of Comparative Content". En Nuccetelli (2003b), pp. 201-218.
- PEACOCKE, C. (1996): "Our Entitlement to Self-Knowledge: Entitlement, Self-knowledge and Conceptual Redeployment". En Ludlow/Martin (1998), pp. 265-303.
- PÉREZ OTERO, M. (2009): "Transmisión de la Justificación Epistémica". En Quesada (2009), pp. 148-174.
- PÉREZ OTERO, M. (2010): "Defensa de una Versión del Argumento Trascendental de Putnam contra la Posibilidad Epistémica de que seamos Cerebros en Cubetas". En Jaume/Liz/Ponte/Vázquez (2010), pp. 139-140.
- PÉREZ OTERO, M. (2011): "Modest Skepticism and Question Begging Proper". *Grazer Philosophische Studien*, 83, pp. 9-32.
- PÉREZ OTERO, M. (2012): "Los Propósitos de Razonar, Ilustrados con el Argumento Externista Anti-Escéptico de Putnam". *Theoria*, 27, pp. 55-74.
- PERRY, J. (1988): "Cognitive Significance and New Theories of Reference". En Perry (2001), pp. 227-247.
- PERRY, J. (2001): *The Problem of the Essential Indexical and other Essays*. Stanford: CSLI.
- PESSIN, A. Y S. GOLDBERG (eds.) (1996): *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning'"*. Londres: Sharpe.
- PLATÓN (2011): *Diálogos I*. Madrid: Gredos

- PRESTON, J. M. (ed.) (1998): *Thought and Language*. Cambridge: Cambridge University Press.
- PRYOR, J. (2001): "Is Moore's Argument an Example of Transmission-Failure?". Conferencia en Pacific Division APA; Seattle.
- PRYOR, J. (2004): "What's Wrong with Moore's Argument". *Philosophical Issues*, 14, pp. 349-378.
- PUTNAM, H. (1970): "Is Semantics Possible?". En Putnam (1975a), pp. 139-152.
- PUTNAM, H. (1973a): "Explanation and Reference". En Putnam (1975a), pp. 196-214.
- PUTNAM, H. (1973b): "Meaning and Reference". En Schwartz (1990), pp. 119-132.
- PUTNAM, H. (1974): "Comment on Wilfrid Sellars". *Synthese*, 27, pp. 445-455.
- PUTNAM, H. (1975a): *Mind, Language and Reality. Philosophical Papers*, vol. 2. Cambridge: Cambridge University Press.
- PUTNAM, H. (1975b): "Language and Philosophy". En Putnam (1975a), pp. 1-32.
- PUTNAM, H. (1975c): "The Meaning of 'Meaning'". En Putnam (1975a), pp. 215-271.
- PUTNAM, H. (1981): *Reason, Truth and History*. Cambridge: Cambridge University Press (Trad. Cast.: en Putnam, H., *Razón, verdad e historia*, Madrid: Tecnos, 1988).
- PUTNAM, H. (1983a): "Possibility and Necessity". En Putnam (1983b), pp. 46-68.
- PUTNAM, H. (1983b): *Realism and Reason*. Cambridge: Cambridge University Press.
- PUTNAM, H. (1988): *Representation and Reality*. Cambridge Mass.: MIT (Trad. Cast.: en Putnam, H., *Representación y Realidad: un Balance Crítico del Funcionalismo*, Barcelona, Gedisa, 2000).
- PUTNAM, H. (1990): *Realism with a Human Face*. Cambridge: Harvard University Press.
- PUTNAM, H. (1992): "Why Functionalism didn't Work". En Putnam (1994), pp. 441-459.
- PUTNAM, H. (1994): *Words and Life*. Cambridge: Harvard University Press.
- QUESADA, D. (ed.) (2009): *Cuestiones de Teoría del Conocimiento*. Madrid: Tecnos.
- QUINE, W. V. O. (1960): *Word and object*. Cambridge, Mass.: MIT Press.
- RÁBADE, S. (1971): *Descartes y la Gnoseología Moderna*. Madrid: G. del Toro.
- RAKOVA, M. (2006): *Philosophy of Mind A-Z*. Edimburgo: Edinburgh University Press.
- RAUTI, A. (2010): "How Use Theories of Meaning Can Accommodate Shared Meanings: A Modal Account of Semantic Deference". *Australasian Journal of Philosophy*, 88, pp. 285-303.
- RICHARD, M. (1990): *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Nueva York: Cambridge University Press.
- ROWLANDS, M. (2003): *Externalism: Putting Mind and World Back Together Again*. Chesham: Acumen.
- RUDD, A. (1997): "Two Types of Externalism". *The Philosophical Quarterly*, 47, pp. 501-507.
- RUSSELL, B. (1905): "On Denoting". En Russell (1956), pp. 39-56.
- RUSSELL, B. (1911): "Knowledge by Acquaintance and Knowledge by Description". En Russell (2009), pp. 191-198.
- RUSSELL, B. (1912): *The Problems of Philosophy*. Londres: Williams and Norgate.
- RUSSELL, B. (1914a): "On the Nature of Acquaintance". En Russell (1956), pp. 127-174.
- RUSSELL, B. (1914b): *Our Knowledge of the External World*. Nueva York: Routledge.
- RUSSELL, B. (1956): *Logic and Knowledge: Essays 1901-1950*. Londres: George Allen and Unwin.
- RUSSELL, B. (1959): *My Philosophical Development*. Nueva York: Routledge.
- RUSSELL, B. (2009): *The Basic Writings of Bertrand Russell*. Nueva York: Routledge.

- RUSSELL, G. (2010): "Analyticity in Externalist Languages". En Sawyer (2010), pp. 186-205.
- RYLE, G. (1949): *The Concept of Mind*. Londres: Hutchinson. (Trad. cast.: *El concepto de lo mental*, Barcelona, Paidós, 1967).
- SALMON, N. (1986): *Frege's Puzzle*. Londres: The MIT Press.
- SALMON, N. (2003): "Are General Terms Rigid?". En Salmon (2007), pp. 100-112.
- SALMON, N. (2005a): *Reference and Essence*. Amherst: Prometheus Books; 2ª ed. ampliada; 1ª ed., Princeton, Princeton University Press, 1981.
- SALMON, N. (2007): *Content, Cognition and Communication : Philosophical Papers II*. Oxford : Oxford University Press.
- SALMON, N. Y S. SOAMES (eds.) (1988): *Propositions and Attitudes*. Oxford: Oxford University Press.
- SAWYER, S. (1998): "Privileged Access to the World". *Australasian Journal of Philosophy*, 76, pp. 523-533.
- SAWYER, S. (1999): "An Externalist Account of Introspective Knowledge". *Pacific Philosophical Quarterly*, 80, pp. 358-378.
- SAWYER, S. (2007): "There is no Viable Notion of Narrow Content". En McLaughlin/Cohen (2007), pp. 20-34.
- SAWYER, S. (ed.) (2010): *New Waves in Philosophy of Language*. Londres: MacMillan.
- SCHANTZ, R. (2004): *The Externalist Challenge*. Berlín: Walter De Gruyter.
- SCHWARTZ, S. (ed.) (1990): *Naming, Necessity and Natural Kinds*. Ithaca: Cornell University Press.
- SEARLE, J. R. (1983): *Intentionality: An Essay in the Philosophy of Mind*. Oxford: Cambridge University Press.
- SEGAL, G. M. (2000): *A Slim Book about Narrow Content*. Londres: The MIT press.
- SHOEMAKER, S. (1994): "Self-Knowledge and Inner Sense". *Philosophy and Phenomenological Research*, 54, pp. 249-314.
- SMITH, Q. (1995): "Marcus, Kripke and the Origin of the New Theory of Reference". *Synthese*, 104, pp. 179-189.
- SOAMES, S. (2002): *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford: Oxford University Press.
- SOBER, E. (2000): *Philosophy of Biology*. Boulder: Westview Press.
- SOSA, E. Y J. KIM (eds.) (2000): *Epistemology: an Anthology*. Malden: Blackwell.
- STEUP, M. Y E. SOSA (eds.) (2005): *Contemporary Debates of Epistemology*. Malden: Blackwell
- STICH S. Y T. A. WARFIELD (eds.) (2003): *The Blackwell Guide to Philosophy of Mind*. Malden: Blackwell.
- STRAWSON, P. F. (1950): "On Referring". *Mind*, 59, pp. 320-344.
- STRAWSON, P. F. (1979): *Individuals: an Essay in Descriptive Metaphysics*. Londres: Methuen
- TALMAGE, C. J. L. (1996): "Davidson and Humpty Dumpty". *Noûs*, 30, pp. 537-544.
- TARSKI, A. (1935): "Die Wahrheitsbegriff in den formalisierten Sprachen". *Studia Philosophica*, 1, pp. 261-405.
- TARSKI, A. (1944): "The Semantic Conception of Truth and the Foundations of Semantics". En Martinich (1996), pp. 61-84. (Trad. cast.: en L.M. Valdés (ed.), *La Búsqueda del Significado*, Madrid, Tecnos, 1991, pp. 299-335).
- TICHÝ, P. (1983): "Kripke on Necessity *A posteriori*". *Philosophical Studies*, 43, pp. 225-241.
- TYE, M. (1998): "Externalism and Memory". *Proceedings of the Aristotelian Society*, 72, pp. 77-94.

- VAN BRAKEL, J. (2000): *Philosophy of Chemistry: Between the Manifest and Scientific Image*. Leuven: Leuven University Press.
- VILANOVA ARIAS, J. (2009): "The Many Faces of the McKinsey's Paradox". En Fernández Moreno (2009), pp. 231-261.
- WARFIELD, T. A. (1992): "Privileged Self-Knowledge and Externalism are Compatible". En Ludlow/Martin (1998), pp. 215-221.
- WARFIELD, T. A. (1997): "Externalism, Self-Knowledge, and the Irrelevance of Slow Switching". En Ludlow/Martin (1998), pp. 231-234.
- WARFIELD, T.A. (1998): "A *a priori* Knowledge of the World: Knowing the World by Knowing our Minds". *Philosophical Studies*, 92, pp. 127-147.
- WEISBERG, M. (2006): "Water is Not H₂O". En Baird/Scerri/McIntyre (2006), pp. 337-346.
- WETTSTEIN, H. (2004): *The Magic Prism: An Essay in the Philosophy of Language*. Oxford: Oxford University Press.
- WIKFORSS, Å.M. (2010): "Are Natural Kind Terms Special?". En Beebe/Sabbarton-Leary (2010b), pp. 64-83.
- WILSON, R. (1999b): "Realism, Essence, and Kind: Resuscitating Species Essentialism?". En Wilson (1999a), pp. 187-207.
- WILSON, R. (ed.) (1999a): *Species: New Interdisciplinary Essays*. Cambridge Mass.: MIT Press.
- WILSON, R. A. (2003): "Individualism" en Stich/ Warfield (2003), pp. 256-287.
- WILSON, R., M. J. BARKER Y I. BRIGANDT (2007): "When Traditional Essentialism Fails: Biological Natural Kinds". *Philosophical Topics*, 35, pp. 189-215.
- WITTGENSTEIN, L. (1967): *Zettel*. Oxford: Blackwell.
- WOODFIELD, A. (1984): *Thought and Object: Essays on Intentionality*. Oxford: Clarendon Press.
- WOODFIELD, A. (1998): "Social Externalism and Conceptual Diversity". En Preston (1998), pp. 77-102.
- WRIGHT, C. (1992): "On Putnam's Proof that We Are Not Brains in a Vat". En Clark/Hale (1994), pp. 216-241.
- WRIGHT, C. (2000a): *Self-Knowledge: The Wittgensteinian Legacy*. En Wright/Smith/Macdonald (2000), pp. 13-45.
- WRIGHT, C. (2000b): "Cogency and Question-begging: Some Reflections on McKinsey's Paradox and Putnam's Proof". *Philosophical Issues*, 10, pp. 140-163.
- WRIGHT, C. (2002): "(Anti-)Sceptics Simple and Subtle: G. E. Moore and John McDowell". *Philosophy and Phenomenological Research*, 68, pp. 330-348.
- WRIGHT, C. (2003): "Some Reflections on the Acquisition of Warrant by Inference". En Nuccetelli (2003b), pp. 57-79.
- WRIGHT, C., B. C. SMITH Y C. MACDONALD (eds.) (2000): *Knowing our Own Minds*. Nueva York: Oxford University Press.
- ZEMACH, E. (1976): "Putnam's Theory on the Reference of Substance Terms". *Journal of Philosophy*, 73, pp. 116-127.

