UNIVERSITAT DE BARCELONA

Auto-conocimiento, memoria y racionalidad
(Estudio de tres argumentos anti-externistas)

Ekain Garmendia Mugica

AUTO-CONOCIMIENTO, MEMORIA Y RACIONALIDAD

(ESTUDIO DE TRES ARGUMENTOS ANTI-EXTERNISTAS)

DOCTORANDO: EKAIN GARMENDIA MUGICA

DIRECTOR DE LA TESIS: MANUEL PÉREZ OTERO
UNIVERSITAT DE BARCELONA, FACULTAT DE FILOSOFIA
DEPARTAMENTO DE LÒGICA, HISTÒRIA I FILOSOFIA DE LA CIÈNCIA
PROGRAMA DE DOCTORADO: CIENCIA COGNITIVA Y LENGUAJE (BIENIO 2004-2006)

ÍNDICE

Parte II

Memoria, Capacidad Conceptual, y Auto-Conocimiento

Parte III

Externismo, Inferencia, y Racionalidad

(PREFACIO)

El objetivo de este trabajo es estudiar tres argumentos anti-externistas. Ninguno consigue su objetivo: demostrar que el modelo externista de la mente tiene algún tipo de problema que no puede esquivar (o ésa será al menos la conclusión a la que llegaremos). Pero casi más que el veredicto sobre el éxito o fracaso de cada uno de los argumentos nos interesan las discusiones que se originan en el camino, las tesis y los compromisos que salen a la superficie, las perspectivas y los modelos que vamos dibujando en cuanto nos adentramos en todas estas cuestiones. De todo ello trata este trabajo.

Guardamos actitudes proposicionales, estados mentales intencionales. Uno cree que hay casas de color rosa y pueblos con menos de dos mil habitantes, desea que el tren de vuelta salga a la hora, teme que amanezca pronto, y juzga que el exilio es el único modo de abolir el destino. Estos estados tan comunes han dado un material ingente a la Filosofía del siglo XX y, a comienzos del XXI, la situación no tiene visos de cambiar. Los estados mentales intencionales, las actitudes proposicionales, han suscitado preguntas en la Filosofía del Lenguaje (¿cuál es la semántica y la pragmática de las adscripciones de estas actitudes?), en Epistemología (¿qué principios sobre evidencia y justificación rigen nuestro conocimiento sobre qué creemos, deseamos o juzgamos?), en la Filosofía de la Psicología (¿supone una psicología popular constituida por este tipo de adscripciones una buena base para una psicología científica?) o la Teoría de la Acción

(¿cuál es el rol explicativo de la adscripción de una creencia en la explicación y justificación de una acción?).

El externismo semántico ofrece una respuesta a una pregunta formulada en la Filosofía de la Mente: una vez fijados los detalles de la situación en que se encuentra un sujeto (una vez determinamos en qué estado cerebral se encuentra, qué propiedades fenoménicas tiene y cuáles son sus disposiciones, su historia causal, cómo es exactamente el entorno que le rodea), ¿qué factores son relevantes para determinar en qué estado mental (intencional) se encuentra ese sujeto?

O, dicho de otra manera: ¿qué determina qué pensamientos y conceptos tiene uno, qué piensa, cree, juzga, teme o desea? El internista responde que sólo los factores *internos* al sujeto, lo que sucede "de piel para adentro", normalmente sus propiedades cerebrales o fenoménicas, son relevantes para fijar sus estados mentales intencionales. Los factores *externos* no son relevantes para determinar qué piensa o teme un sujeto—si mantienes fijas sus propiedades internas, mantienes fijas sus propiedades mentales, da igual qué cambios introduzcas en su entorno. El externista rechaza este modelo: hay elementos externos al sujeto que en parte determinan qué conceptos tiene éste, qué cree, piensa o teme.

Tan pronto comenzó a extenderse este modelo externista (allá por la década de 1970), surgieron los primeros problemas y objeciones. Desde el comienzo se generalizó la sospecha de que el externismo semántico mostraba problemas de compatibilidad con la tesis del auto-conocimiento autoritativo (la idea de que, en un gran número de casos, podemos saber qué pensamos, creemos o juzgamos sin necesidad de estudiar cómo es nuestro entorno)—si qué estoy pensando depende de factores externos, ¿cómo puedo saber qué pienso sin saber primero cuáles son esos factores externos? La sospecha se hizo argumento, y sobre este argumento tratará la primera parte del trabajo.

Pero digamos un par de cosas sobre terminología y notación antes de dar un primer vistazo a los contenidos del trabajo. La bibliografía está en inglés, este texto en castellano, muchas veces hemos tenido que tomar decisiones sobre cómo traducir este término o aquel concepto—esperamos que las malas decisiones no dificulten la lectura del trabajo. Querríamos avisar sobre un par de decisiones terminológicas. Nos hemos

decantado por *internismo y externismo semántico* en detrimento de *individualismo y anti-individualismo*. La decisión es poco menos que arbitraria; no creemos que haya motivos importantes que favorezcan el uso de unos términos más que el de otros, y nos hemos alineado con quienes creemos que son mayoría. Por otro lado, cuando tratemos temas relacionados con el auto-conocimiento, hablaremos de *conocimiento basado en introspección*, y no *conocimiento a priori*. El motivo es simple (aunque tampoco muy importante): nos parece que "S sabe *a priori* que *p*" típicamente implica que "*p* es cognoscible *a priori*", pero que yo esté pensando que *p* no es cognoscible "*a priori*" para nadie más que yo. Por eso, nos parece que el uso de '*a priori*' en este contexto concreto "chirría". Además, usamos más *auto-conocimiento autoritativo* que *acceso privilegiado*, aunque entendemos que estos términos son sinónimos, y en algunos momentos también usaremos el segundo.

Hay autores que no comparten estos usos con nosotros (concretamente estamos pensando en Jessica Brown y Tyler Burge), cuando los citemos por supuesto respetaremos el original, y es posible que cuando presentemos y discutamos sus opiniones no nos ciñamos estrictamente a un uso u otro. Sea como fuere, usaremos 'externismo' y 'anti-individualismo' como sinónimos y, cuando tratemos cuestiones de auto-conocimiento, también lo haremos así con 'conocimiento mediante introspección' y 'conocimiento *a priori*'.

En cuanto a la notación, digamos sólo que usaremos la comilla simple cuando hablemos sobre términos lingüísticos y enunciados, y caracteres versales cuando lo hagamos sobre conceptos y pensamientos (París es una ciudad, 'París' una palabra, PARÍS un concepto).

Mencionemos también cuáles han sido las prácticas seguidas en la elaboración del trabajo (la "metodología"). Tratándose de un trabajo de investigación filosófica, creemos que podemos identificar tres pasos principales: estudio de las fuentes, elaboración de las hipótesis, y contrastación de éstas (mediante discusión en diversos seminarios o presentaciones).

El trabajo tiene tres partes, cada una sobre un argumento anti-externista, y un capítulo introductorio—además de un último capítulo con las conclusiones. Los argumentos estudiados en la primera y segunda parte son *incompatibilistas* (a saber, aspiran a

demostrar que el externismo semántico es incompatible con la tesis del auto-conocimiento autoritativo), en la tercera parte nos centraremos en un argumento que pretende demostrar que el externismo semántico mina nuestras capacidades lógicas (y con ellas, las prácticas racionalizadoras que sustentan).

Comenzaremos el trabajo con un capítulo introductorio, cuyo objetivo será proponer un esbozo del modelo externista. No hilaremos fino, lo que haremos será poco más que bosquejar el posicionamiento al que van dirigidos los argumentos que son el objeto de este trabajo, el lugar desde donde se les responderá. Con este fin, diferenciaremos entre tres tipos de externismo (externismo de clases naturales, externismo social y externismo singular) dependiendo de qué factores externos se juzgue que pueden ser relevantes a la hora de individuar los conceptos y las actitudes proposicionales de uno, y repasaremos (aunque sea por encima) las opiniones de tres autores ya clásicos, cruciales en el desarrollo de esta teoría—Hilary Putnam, Tyler Burge y Gareth Evans. Identificaremos cuál es la tesis definitoria del externismo semántico, enumeraremos algunas de las tesis de Filosofía de la Mente y del Lenguaje a las que típicamente se adhiere el defensor de este modelo, y veremos que la tesis externista parece estar en cierta tensión con la tesis de la transparencia del contenido (la idea de que siempre podemos saber si dos de nuestros pensamientos o conceptos son el mismo o no). Como veremos en este trabajo, el supuesto compromiso del externista a negar que el contenido es transparente está en el origen de varios argumentos anti-externistas.

En la primera parte presentaremos y discutiremos el que seguramente es el argumento incompatibilista más popular. Comenzaremos describiendo un ejemplo de *transición lenta*[1], estos es, un ejemplo donde un individuo cambia inadvertidamente de entorno (cada uno de los argumentos que estudiaremos se sustenta en un ejemplo de este tipo, cada uno con sus propias peculiaridades). Presentaremos dos versiones del argumento, la primera debida a Paul Boghossian (los otros dos argumentos anti-externistas que estudiaremos en la segunda y tercera parte también son suyos), la segunda a Jessica Brown. El argumento viene a decir que, si el externismo es verdadero, entonces puede darse el caso de que la evidencia introspectiva que obtiene un sujeto será compatible con un escenario relevante en el que está pensando algo distinto (Boghossian), o que no

[1] Es cómo hemos traducido el *slow switch* del inglés.

será capaz de discriminar entre su escenario actual y ese escenario relevante (Brown), y que en esas situaciones no estará en posición de saber mediante introspección qué está pensando. Haremos explícitos los dos principios sobre evidencia, justificación y conocimiento detrás de esas dos versiones del argumento (los conoceremos como (RA) y (Discp)), y veremos qué relación guardan entre sí esos dos principios.

Nos centraremos en tres respuestas: la de Burge, la de Falvey y Owens, y la de McLaughlin y Tye. También veremos que Brown plantea su propia respuesta, que arguye que los escenarios contrafácticos que podrían minar nuestro auto-conocimiento no son normalmente relevantes, y veremos cómo podría plantear uno esta misma respuesta partiendo de la idea de que, en las transiciones lentas, uno *reemplaza* su antiguo concepto por aquel que recibe. Pero concluiremos que negar la relevancia de los escenarios alternativos no constituye una alternativa aceptable, porque esta respuesta entiende mal la dialéctica de la discusión, y porque no dice nada sobre algunos escenarios que podrían minar nuestro conocimiento de pensamientos perceptivos basados en ostensión.

Nos basaremos en las ideas de Falvey y Owens, por un lado, y las de McLaughlin y Tye, por el otro, para bosquejar la que creemos es la mejor respuesta al argumento incompatibilista. Falvey y Owens identifican el principio epistémico en el que se basa el argumento en la versión de Boghossian, (RA), y niegan que sea generalmente verdadero—se les critica algunas veces que no demuestran la falsedad del principio, nosotros propondremos un contraejemplo a (RA) y (Discp). Por contra, McLaughlin y Tye argumentan que el externista puede aferrarse a principios de ese tipo, pero que el argumento incompatibilista no supone ninguna amenaza para él, ya que descansa sobre una noción equivocada de la evidencia introspectiva. Sobre esta base, nosotros mantendremos que el defensor del argumento incompatibilista se encontrará ante *una disyunción incómoda*: o bien niega que los principios epistémicos como (RA) y (Discp) son generalmente verdaderos, o bien rechaza que la evidencia se individúa internamente—ambas tesis no pueden ser verdaderas a la vez.

El problema es que el argumento necesita de esas dos tesis. Por eso, no resulta una amenaza seria para el externista y, por eso, el modelo observacional del auto-conocimiento sobre el que se sustenta tampoco puede ser verdadero. Además, siguiendo

la línea marcada por Falvey y Owens, también nosotros apostaremos por negar la validez de esos principios epistémicos en el campo del auto-conocimiento.

En la segunda parte estudiaremos *el argumento de la memoria* de Boghossian que, como el primero, pretende demostrar que el externismo semántico es incompatible con el auto-conocimiento autoritativo. El argumento de la memoria explota una supuesta consecuencia del externismo: que la víctima de una transición lenta no podría recordar qué pensó en algún momento pasado. Pero esto, se supone, es un problema, ya que podemos estipular que el sujeto no olvidó nada y, según Boghossian, hay dos explicaciones posibles a que alguien no sepa algo en un momento dado: o bien lo ha olvidado, o bien nunca lo supo. Por eso, la víctima de la transición nunca supo lo que pensó, tampoco cuando todavía no había cambiado de entorno.

Pero ya desde un principio veremos que el argumento no demuestra aquello que pretende, porque, como han señalado distintos autores, sus premisas se basan en un uso ambiguo de 'olvidar'. Una vez desambiguamos ese uso, vemos que al menos una de las premisas del argumento habrá de ser falsa—no entraremos a discutir cuál de las dos lo es, ya que opinamos que esta discusión es por completo terminológica (y, por ello, insustancial).

Pero lo que sí nos interesa del argumento es que abre la puerta a un debate que concierne a los compromisos del externista sobre la memoria; prácticamente toda la segunda parte trata de estos compromisos. Primero, veremos cómo hay quien niega que el externismo semántico tiene las consecuencias sobre la memoria que asume Boghossian que tiene, así como quien acepta que la víctima de una transición lenta perdería gran parte del conocimiento que tenía. Los primeros mantienen que cuando la víctima de la transición adquiere el nuevo concepto, éste *cohabita* con sus conceptos anteriores; los segundos que el nuevo concepto *reemplaza* otro anterior.

Seguiremos discutiendo una variante del ejemplo, introducida por John Gibbons, responderemos a un argumento que ofrece éste sobre la base de esa nueva variante, y aprovecharemos este nuevo escenario y esta discusión para explicitar las predicciones a las que pretendemos llegar sobre qué puede recordar y qué no en qué condiciones la víctima de una transición. Una vez identificado a dónde queremos llegar,

bosquejaremos lo que llamamos una *(proto)teoría de la memoria*, con la intención de dejar claro qué determina el contenido del recuerdo de uno, y cuáles son las relaciones justificatorias que toman parte en ese recuerdo. Diferenciaremos entre memoria episódica y memoria semántica, y defenderemos que en la base de la primera sí hay un elemento justificatorio de carácter mnemónico, pero no así en la de la segunda. Sobre la base de esta (proto)teoría, llegaremos a las predicciones que identificamos al principio; así describiremos qué puede recordar y qué no la víctima de una transición, concluyendo que cuando sea incapaz de recordar algo, eso no será así por cuestiones relacionadas con la teoría externista, sino porque ha adquirido nueva evidencia que mina la justificación que tuvo antes para recordar.

Terminaremos comparando nuestra propuesta con las posiciones que hemos descrito al comienzo de la segunda parte, concluyendo que la nuestra es preferible. En concreto, veremos que la idea del reemplazo conceptual no es deseable, especialmente por tres motivos principales: no está claro por qué el contacto con un entorno nuevo debería concluir con la pérdida de un concepto anterior, da resultados extraños cuando se combina con la idea de que hay predominio de las transiciones lentas, y no ofrece una caracterización satisfactoria de la memoria episódica.

En la última parte del trabajo discutiremos un tercer argumento introducido por Boghossian, que trata sobre las supuestas consecuencias que tiene el externismo semántico para nuestras adscripciones de creencias y las explicaciones de racionalidad y conducta. El argumento se basa en un ejemplo de transición lenta, donde el sujeto comete una falacia de equivocación—es el supuesto compromiso del externista a negar que el contenido es transparente lo que está en el origen de la equivocación. El problema es que, de acuerdo con Boghossian, esto tiene la consecuencia de que la víctima de la transición no puede saber *a priori* qué inferencia es válida y cuál no; pero si esto es así, afirma Boghossian, la reflexión *a priori* no le basta para llegar a ser racional, y las adscripciones *de dicto* del externista no servirán para racionalizar sus creencias y explicar su conducta.

Una posible respuesta viene a ser intentar compatibilizar el externismo semántico con la transparencia del contenido. El único modo que vemos de hacer esto es acudiendo al modelo del reemplazo conceptual, pero ya hemos concluido en la segunda parte que el

externista no debería adherirse a este modelo—defenderemos, pues, que debería negar que el contenido es transparente. También presentaremos una estrategia defendida por Stephen Schiffer y Tyler Burge, pero mantendremos que no es acertada.

Veremos que otros muerden la bala lanzada por Boghossian—unos proponen remodelar nuestra noción tradicional de racionalidad, otros asumen que algunas veces el conocimiento *a posteriori* marca alguna diferencia en nuestra racionalidad. Pero argumentaremos que esta posición no es convincente: proponen una alternativa un tanto extrema que el externista no tiene por qué asumir y, además, tenemos dudas sobre si entienden correctamente el argumento que pretenden responder.

Porque, como veremos, el externista tiene a mano una explicación sencilla de los ejemplos que se le suponen problemáticos. Defenderemos que Boghossian interpreta de manera incorrecta los ejemplos en los que se basa su argumento, y que el externista puede explicar a la perfección que los sujetos de estos ejemplos son racionales si acude a algunas de sus creencias de identidad. Arguyendo que en los casos como el presentado por Boghossian hay premisas ocultas sobre relaciones de identidad, podemos concluir que es simplemente falso que el sujeto en cuestión acepta como válida una inferencia que de hecho no lo es. Además, esta respuesta es compatible con negar que el contenido es transparente.

Veremos que nuestra propuesta tiene una consecuencia que a algunos quizás les podrá parecer extraña (pero que nosotros creemos que el externista tendría que aceptar): algunos elementos externos al sujeto (concretamente: la veracidad o falsedad de sus creencias de identidad) son relevantes para determinar cuáles son las inferencias que sigue ese sujeto, para determinar si éstas contienen o no una premisa de identidad—las propiedades internas de un sujeto no bastan para determinar qué premisas contienen algunas de las inferencias que sigue.

Concluiremos que los ejemplos expuestos no suponen ningún problema relevante para el externista, y que el argumento no logra demostrar que éste tenga problemas con las adscripciones *de dicto*, o con racionalizar creencias y explicar conductas.

Creemos necesario notar que, entre los argumentos anti-externistas que estudiaremos, no se encuentra uno que ha generado tanta discusión como bibliografía—nos estamos refiriendo al conocido como "argumento McKinsey" (McKinsey (1991)). Los motivos son dos: por un lado opinamos que los tres argumentos que estudiaremos tienen características comunes de las que el argumento McKinsey carece, por el otro, este trabajo no nos permitía el espacio (y el tiempo) suficiente para poder presentar y estudiar este argumento adecuadamente.

****

Quisiera terminar este prefacio dando las gracias a algunas (de las muchas) personas que han hecho que acabar este trabajo haya sido posible.

Quiero primero agradecer a mi director, Manuel Pérez Otero, sus esfuerzos (algunos de ellos en vano, me temo) para que no me perdiera en cada uno de los embrollos en que me he empeñado en verme metido. Sólo espero que el resultado esté a la altura.

Me he pasado los últimos cinco años (y algo más) entre la gente de LOGOS—cuesta imaginar un entorno mejor para desarrollar un trabajo de filosofía. Manuel García Carpintero, Dan López de Sa, Josep Macià, Genoveva Martí, David Pineda, Josep Lluis Prades, Sven Rosenkranz, Pepa Toribio y los demás miembros del grupo siempre han sido un modelo a seguir (aunque los cafés y las muchísimas discusiones con Adrian, Andrei, Chiara, Dan, Fiora, Jose, Luis, Manolo, Marc, Mirja, Mireia, Oscar, Sebas, Sergi y otros futuros filósofos no han sido menos importantes en mi aprendizaje).

Pasé la primavera del 2008 en el Departamento de Filosofía de la Northwestern University de Evanston (Illinois). Sandy Goldberg fue un excelente anfitrión—este trabajo le debe mucho.

Comencé a interesarme por esto de la Filosofía Analítica en la Universidad del País Vasco, en alguna clase de Agus Arrieta o Kepa Korta. Pero la mayoría de mis recuerdos de licenciatura tienen que ver con Haritz, Ibon, Ibone, Idoia, Leire y Oihana, y a éstos les debo en parte que, aún hoy, relacione 'Filosofía' con algo que no sé muy bien qué es, pero es sin duda ameno y divertido.

Durante cuatro años he recibido una beca predoctoral otorgada por la Consejería de Educación del Gobierno Vasco, que ha sido la fuente principal de financiación de este trabajo. Además, he recibido la ayuda de dos proyectos de investigación: "La constitución del contenido representacional. Aspectos semánticos y epistemológicos" (HUM2005-07539-C02-01; Investigador Principal: M. Pérez Otero), Ministerio de Educación y Ciencia; y "Discriminabilidad: representación, creencia y escepticismo" (FFI2008-06164-C02-01; Investigador Principal: M. Pérez Otero), Ministerio de Ciencia e Innovación; así como del Proyecto CONSOLIDER-INGENIO 2010, "Perspectival Thoughts and Facts" (CSD2009-00056; Coordinador: M. García-Carpintero), Ministerio de Ciencia e Innovación y del Grupo de investigación consolidado LOGOS, Research Group in Logic, Language and Cognition (2009 SGR 1077; Coordinadora: G. Martí), Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya.

Y quiero dar las gracias a Idoia, Maider y Rosa—por los cafés, los cines, las discusiones sobre teoría de género, alguna noche en vela, y sufrirme mañana sí y mañana también (sobre todo los últimos meses, en los que he estado incluso más gruñón que de costumbre).

Aitari eta amari. Joanari.

(eskerrik asko)

(0).....

INTRODUCCIÓN: REPASO HISTÓRICO

Hay un cambio de modelo en la Filosofía de la Mente y La Teoría de la Referencia, que se da entre las décadas de 1960 y 1970. La gran mayoría de autores (si no todos[2]) eran *internistas* hasta el momento; la gran mayoría son *externistas* desde entonces. Algunos han presentado distintos argumentos contra esta teoría externista—el grueso de este trabajo trata sobre tres de esos argumentos.

El objetivo de esta introducción es proponer un esbozo más o menos general de la teoría externista que atacan los tres argumentos que presentaremos después. Nuestra intención es acabar esta introducción con una caracterización de cierto *hipotético externista* que, aunque no se debe identificar con ningún autor en concreto, se adheriría a las tesis más importantes que delimitan este modelo. Para tal fin, expondremos muy brevemente las opiniones mantenidas por tres autores ya clásicos, cruciales en el desarrollo de esta corriente: Hilary Putnam, Tyler Burge y Gareth Evans.

Pero nos es necesario comenzar al menos bosquejando aquel corpus teórico que rechaza el modelo externista en el que estamos interesados. En dos palabras, el internismo semántico es una teoría sobre qué factores determinan cuáles son las propiedades

[2] Es probable que el *segundo* Wittgenstein sea la gran excepción.

intencionales de los estados mentales de un sujeto, sobre qué determina qué conceptos, pensamientos, creencias o deseos tiene un sujeto—y así lo será también el externismo semántico. Aunque, como hemos dicho, la gran mayoría de filósofos del lenguaje y de la mente anteriores a 1970 (más o menos) se puede considerar internista, también es cierto que no son muchos los que explícitamente se adhieren a esta tesis (que definiremos en unos pocos párrafos). A pesar de ello, sí es verdad que es una idea que subyace al modelo de Filosofía del Lenguaje y de la Mente y Teoría de la Referencia más popular hasta ese momento.

Creemos que la siguiente es una definición adecuada del internismo semántico:

> **Internismo semántico:** Las propiedades *internas* de un sujeto (sus estados fenoménicos o cerebrales) bastan para determinar qué conceptos tiene o en qué estado mental intencional[3] se encuentra (qué piensa, juzga, cree o desea)

Aclaremos un poco los conceptos que aparecen en esta definición. Llamamos estados mentales *intencionales* a todos aquellos estados que tienen propiedades representacionales o contenidos, característicamente *actitudes proposicionales*—como pensar que *p*, juzgar que *p*, creer que *p*, o desear que *p*. Estos estados están constituidos por una actitud (pensar, juzgar, creer, desear) y una proposición, aquello hacia lo cual se guarda la actitud. Por ejemplo, la creencia de uno de que la tierra es redonda, su creencia de que hoy es Martes, o de que algunas casas son de color rosa comparten actitud, *creer*, pero difieren en contenido. Por contra, la creencia de uno de que pronto se hará de noche, su deseo de que se haga de noche pronto, o su temor a que pronto se haga de noche difieren en actitud, pero comparten contenido (la proposición de que pronto se hará de noche). Estas proposiciones que en parte constituyen nuestras actitudes proposicionales están constituidas por *conceptos*—uno no puede temer que se haga de noche pronto sin tener el concepto NOCHE o el concepto PRONTO, ni creer que hay casas de color rosa sin tener el concepto ROSA.

[3] Por supuesto, es muy probable que quien defienda un modelo internista para los estados mentales intencionales también lo haga para los estados no-intencionales. No es extraño que uno sea externista sobre estados intencionales pero internista sobre estados no-intencionales, y por eso no discutiremos éstos en nuestro trabajo.

Lo dicho: el internismo viene a decir que qué conceptos tiene un sujeto (y, por lo tanto, qué actitudes proposicionales guarda) depende completamente de sus propiedades internas, de cómo es de piel para adentro—una vez fijado cómo son los adentros de un sujeto, se fija también en qué estado mental se encuentra. Es común identificar estos estados internos con los estados fenoménicos o físicos (estados cerebrales, neuronales). Lo que suceda "más allá del sujeto" no es relevante para determinar qué conceptos tiene, o qué actitudes proposicionales guarda.[4]

Se sigue que si dos sujetos S y S' están en los mismos estados internos, en los mismos estados fenoménicos y cerebrales, entonces se encontrarán en los mismos estados mentales. S tendrá la creencia de que *p* o el deseo de que *q* si y sólo si S' también lo tiene—es imposible que dos sujetos sean internamente indistinguibles y que difieran en sus estados mentales.

> Supón que tengo un *doppelgänger* (...) que es "idéntico" a mí molécula a molécula. Si eres dualista, entonces supón también que mi *doppelgänger* piensa los mismos pensamientos verbalizados que yo, tiene los mismos datos de los sentidos, las mismas disposiciones, etc. Es absurdo pensar que *su* estado psicológico es en lo más mínimo diferente al mío.[5]

Esta tesis al menos subyace a cierto modelo sobre qué determina los contenidos y las referencias de los conceptos que tiene un sujeto y los términos lingüísticos que emplea. En lo que sigue enumeraremos algunas tesis características de este modelo. No pretendemos hilar fino; puede que lo que sigue no sirva para caracterizar a ningún autor en concreto—puede que algún autor característicamente internista rechazara alguna o varias de las tesis que siguen. Pero el esbozo que sigue sí nos servirá para ir presentando el modelo externista en el que estamos interesados.

Este modelo tradicional asume que tenemos cierto tipo de *acceso íntimo* a los contenidos de nuestros estados mentales—a nuestros conceptos. Más concretamente,

---

[4] Por supuesto, el internismo no viene a decir que las relaciones que tiene un sujeto con su entorno exterior *no pueden* influir en que *adquiera* un concepto u otro, una creencia u otra, (por ejemplo, uno típicamente adquirirá el concepto ARAÑA al tener algún tipo de contacto con las arañas, o adoptará la creencia LAS ARAÑAS PUEDEN SER PELUDAS al *ver* que alguna araña lo es). Lo que mantiene el internismo es que el exterior es irrelevante para *determinar* qué conceptos tiene uno (no que lo sea para que los adquiera).

[5] Suppose I have a Doppelgänger (...) who is molecule for molecule "identical" with me. If you are a dualist, then also suppose my Doppelgänger thinks the same verbalized thoughts I do, has the same sense data, the same dispositions, etc. It is absurd to think *his* psychological state is one bit different from mine. (Putnam (1973), p. 309)

asume que las descripciones y creencias que relacionamos con nuestros conceptos determinan por completo su contenido y referencia. Pongamos un ejemplo.

Rosa guarda varias actitudes proposicionales que tienen a Manu como objeto, tiene creencias y deseos sobre Manu. Esto es así porque Rosa tiene un concepto MANU que refiere a Manu. Según el modelo tradicional, esto es posible porque Rosa tiene una descripción ϕ relacionada con ese concepto; por ejemplo, la descripción "El chico moreno con perilla y nariz aguileña que tan bien imita a Rita Barberá". Y, de acuerdo con el modelo tradicional, la descripción "El chico moreno..." nos muestra el contenido de MANU al mismo tiempo que determina cuál es su referencia. Cuando un sujeto tiene un concepto C que relaciona con una descripción ϕ, C y ϕ comparten sentido—el sentido de MANU (su contenido) es el sentido de "El chico moreno..." (su contenido). Además, C referirá al único objeto *x* que de hecho satisfaga la descripción ϕ (cualquiera que sea el objeto *x*)—MANU refiere a Manu, porque éste es el único objeto que satisface la descripción "El chico moreno...".

De acuerdo con este modelo, sabemos qué es lo que determina tanto el contenido de nuestros conceptos como su referencia. Porque es la descripción ϕ que relaciona un sujeto S con su concepto C lo que determina su contenido y referencia—si S entendiera que la descripción que define el concepto en cuestión no es ϕ, sino φ (que difiere en contenido y referencia con ϕ), entonces el concepto que estaría identificando sería otro (con un contenido y una referencia distintos); y nos es por completo conocido cuáles son estas descripciones que relacionamos con nuestros conceptos.

Estas afirmaciones sobre qué determina el contenido y la referencia de un concepto encuentran su equivalente lingüístico. Así, como dice Putnam (1975),

> La teoría del significado vino a descansar sobre dos asunciones indiscutidas:
>
> (I) Que conocer el significado de un término es sólo una cuestión de estar en cierto estado psicológico (...)
> (II) Que el significado de un término (en el sentido de 'intensión') determina su extensión (en el sentido de que la mismidad de intensión implica mismidad de extensión).[6]

[6] The theory of meaning came to rest on two unchallenged assumptions: (I) That knowing the meaning of a term is just a matter of being in a certain psychological state (...) (II) That the meaning of a term (in the

Cuando usamos un término t, expresamos un concepto C. Así, si lo que determina el contenido y la referencia de C es cierta descripción ϕ, la misma descripción determinará el contenido y la referencia de t. El término 'Manu' que usa Rosa tiene el mismo sentido que la descripción "El chico moreno...", y referirá al único objeto *x* que satisfaga esa descripción, a Manu.

Debido a estas cuestiones, este modelo tradicional ha asumido cierta noción concreta de competencia lingüística. Si S expresa el concepto C al usar el término t, y la descripción ϕ determina el contenido de ambos, de C y de t, entonces ϕ será la *definición* de t (y de C). Así, la competencia lingüística viene a ser saber cuál es la descripción ϕ que define el término t, muestra su sentido y determina su referencia—una cuestión de relacionar correctamente cada término t con el correspondiente concepto C. Uno comprende y usa un término competentemente si y sólo si sabe cuál es la descripción que determina su contenido y su referencia—la descripción que define el término. Por eso, la comprensión que tiene un hablante competente de un término de su lenguaje es *completa*; el hablante sabe cuál es la definición del término que delimita su extensión.

Esto es, según el modelo tradicional, es imposible que uno no sepa cuál es la descripción que delimita el sentido de uno de sus conceptos, o que no sepa qué condiciones ha de cumplir algo para caer dentro de la extensión de uno de sus conceptos, y, también según este modelo tradicional, en cuanto un hablante S es un hablante *competente* del término t, sabe cuál es la definición de t, cuál es su referencia, o qué condición ha de cumplir algo para caer dentro de su extensión.

Terminemos este bosquejo del internismo semántico y del modelo sobre determinación del contenido y la referencia al que subyace presentando una tesis que tradicionalmente se ha asumido como verdadera—que el contenido es *transparente*:

> Con respecto a cualesquiera dos de sus pensamientos o creencias, un individuo puede conocer autoritativa y directamente (esto es, sin basarse en inferencias desde su entorno observado) si tienen el mismo contenido o no.[7]

---

sense of 'intension') determines its extension (in the sense that sameness of intension entails sameness of extension). (Putnam (1975), p. 218)

[7] With respect to any two of his thoughts or beliefs, an individual can know authoritatively and directly (that is, without relying on inferences from his observed environment) whether or not they have the same content. (Falvey y Owens (1994), pp. 109-110)

O, lo que sería su equivalente lingüístico:

> Es una característica innegable de la noción de significado—siendo esa noción tan oscura como es—que el significado es *transparente* en el sentido de que, si uno dona significado a cada una de dos palabras, tiene que saber si esos significados son el mismo.[8]

Esto es, brevemente, sean C y C' dos conceptos que tiene S, y *p* y *q* dos de sus pensamientos. Según la tesis de la transparencia del contenido, S puede saber mediante introspección si C y C', por un lado, y *p* y *q*, por el otro, tienen el mismo contenido o no. Además, si S es un hablante competente de los términos t y t' y de los enunciados s y s', entonces sabe si t y t' tienen el mismo contenido o no, y si s y s' expresan la misma proposición o no[9]. Podemos reducir la afirmación de que el contenido es transparente a las siguientes dos tesis (cuando a lo largo del trabajo tratemos sobre la transparencia del contenido, será la siguiente formulación la que tendremos en mente):

> **Transparencia de mismidad de contenido:** para cualesquiera dos pensamientos, o constituyentes de pensamiento, que *S* considera en el momento *t*, si tienen el mismo contenido, entonces, en *t*, *S* puede darse cuenta a priori de que tienen el mismo contenido.
> **Transparencia de diferencia de contenido:** para cualesquiera dos pensamientos, o constituyentes de pensamiento, que *S* considera en el momento *t*, si tienen contenidos diferentes, entonces, en *t*, *S* puede darse cuenta a priori de que tienen contenidos diferentes.[10]

Terminemos estos primeros párrafos preliminares. Acabamos de esbozar un modelo de Filosofía de la Mente y del Lenguaje y de Teoría de la Referencia que, hemos dicho, descansa sobre la tesis internista—a saber, que las propiedades internas de un sujeto

---

[8] It is an undeniable feature of the notion of meaning—obscure as that notion is—that meaning is *transparent* in the sense that, if someone attaches a meaning to each of two words, he must know whether these meanings are the same. (Dummett (1978), p. 131)

[9] Ya Frege defendía que el contenido es transparente: "¿Cómo es posible (...) que fuera dudoso si un signo simple tiene el mismo sentido que una expresión compleja, si conocemos no sólo el sentido del signo simple, sino que también podemos reconocer el sentido de la expresión compleja por el modo en el que está compuesta? La cuestión es que si realmente tenemos una comprensión clara del sentido del signo simple, entonces no puede ser dudoso si coincide con el sentido de la expresión compleja" ("How is it possible (...) that it should be doubtful whether a simple sign has the same sense as a complex expression, if we know not only the sense of the simple sign, but can recognize the sense of the complex one from the way it is put together? The fact is that if we really do have a clear grasp of the sense of the simple sign, then it cannot be doubtful whether it agrees with the sense of the complex expression" (Frege (1914), p. 211)

[10] **Transparency of sameness of content:** for any two thoughts, or thought constituents, that *S* entertains at time *t*, if they have the same content then, at *t*, *S* can realize a priori that they have the same content.
Transparency of difference of content: for any two thoughts, or thought constituents, that *S* entertains at time *t*, if they have different contents then, at *t*, *S* can realize a priori that they have different contents. (Brown (2004), p. 160)

bastan para determinar sus estados mentales intencionales (qué conceptos tiene, que pensamientos piensa, qué creencias cree). El externista rechaza esta tesis, y con ella (al menos algunas de) las afirmaciones que hemos mencionado en los párrafos anteriores; esto es, el externista afirma que las propiedades internas de un sujeto no bastan para determinar en qué estado mental se encuentra, que algunos factores externos a él pueden ser relevantes a la hora de concluir qué conceptos tiene o qué pensamientos piensa.

Pero hay distintas formas de desarrollar esta idea, dependiendo de cuáles crea uno que son los factores externos relevantes a la hora de identificar los estados mentales de un sujeto. Así, y basándonos en las opiniones de tres autores ya clásicos (Putnam, Burge y Evans), en las siguientes secciones presentaremos tres versiones del externismo semántico: externismo de clases naturales, externismo social y externismo singular. Además, aprovecharemos las opiniones de Evans (1982) para explicar cómo es posible compatibilizar las ideas externistas con cierto modelo neo-fregeano sobre la naturaleza del contenido.

0.1. HILARY PUTNAM: EXTERNISMO DE CLASES Y DIVISIÓN DEL TRABAJO LINGÜÍSTICO

Hilary Putnam publicó "The Meaning of 'Meaning'" en 1975[11]. Esboza en él un modelo teórico sobre la semántica y teoría de la referencia de los términos de clase natural que choca directamente con el modelo tradicional que hemos descrito en los párrafos anteriores. Entre otras cosas, afirma que las propiedades superficiales características de una clase no bastan para determinar el contenido y la extensión del correspondiente término, que nuestra competencia lingüística y la determinación de la referencia reside en parte en una "división del trabajo lingüístico", o que a una clase natural le es esencial cierta "estructura oculta" que generalmente nos descubre la ciencia.

[11] Y "Meaning and Reference" en 1973, donde ya adelanta gran parte de las tesis defendidas en "The Meaning of 'Meaning'". Pero éste último es bastante más extenso que el primero y desarrolla en él algunas afirmaciones hechas en "Meaning an Reference"; por eso nos centraremos nosotros en él.

En el mismo artículo Putnam reivindica que "los significados no están en la cabeza"; esta formulación ha tenido tanto eco que podríamos caracterizarla como uno de los lemas externistas más populares. Pero ya desde el comienzo queremos dejar clara una cosa. Putnam (1975) limita sus propuestas al significado lingüístico, a las propiedades representacionales o de contenido *de los términos lingüísticos*. No es externista sobre los estados mentales, acepta que si dos individuos están en la misma situación interna, entonces sus estados mentales tendrán las mismas propiedades de contenido. Aún así, dada la repercusión de las tesis presentadas en "The Meaning of 'Meaning'", creemos oportuno exponer brevemente estas ideas para, basándonos en ellas, moldear el punto de vista de un hipotético externista de clases. A ello dedicamos esta sección.

Putnam (1975) comienza por presentar dos principios que, según él, le son esenciales al modelo semántico tradicional (ya los hemos expuesto al comienzo de este capítulo introductorio[12]). El primero dice que conocer el significado de un término, su *intensión* o *contenido*, no es más que estar en un estado mental concreto—los estados mentales de un sujeto bastan para determinar los significados de los términos que emplea; el segundo que el *contenido* o *intensión* de un término determina su extensión. Puede haber dos términos que comparten extensión pero que difieren en intensión, no así dos términos que comparten intensión pero no extensión. Dado que Putnam asume que los estados internos de un sujeto determinan sus estados mentales, se sigue que las propiedades internas de uno bastan para determinar la extensión de los términos que emplea.

Pero esto es falso, "La extensión no queda determinada por el estado psicológico" (Putnam (1975), p. 353). Al menos uno de los dos principios (I) y (II) que en parte caracterizan el modelo tradicional habrá de ser falso. Para probar tal cosa, Putnam introduce los experimentos mentales de Tierra Gemela, los cuales han dado pie a varios experimentos y argumentos filosóficos—los "ejemplos de transición lenta" que discutiremos extensamente a lo largo de este trabajo también se basan en este tipo de experimentos. Describiremos aquí el más famoso de ellos, aunque sin entrar en demasiados detalles.

[12] (I) Que conocer el significado de un término es sólo una cuestión de estar en cierto estado psicológico (...) (II) Que el significado de un término (en el sentido de 'intensión') determina su extensión (en el sentido de que la mismidad de intensión implica mismidad de extensión).

Supongamos que existe un planeta, al que nosotros llamaremos 'Tierra Gemela'. La Tierra Gemela es muy parecida a la Tierra: está habitada por humanos, quienes llevan un estilo de vida muy parecido al que llevamos nosotros a comienzos del siglo XXI, y comparte con la Tierra su flora y su fauna[13]. Además, los habitantes de la Tierra Gemela hablan castellano (o, al menos, algo muy parecido). Pero hay una pequeña diferencia entre la Tierra y la Tierra Gemela. Aquello que los habitantes de la Tierra Gemela llaman 'agua', el líquido transparente e insípido que beben, no es H_2O—tiene una estructura química compleja que nosotros abreviaremos como XYZ. Comparte sus propiedades superficiales con el agua (hierve a cien grados y se hiela a cero, es transparente e insípida, cae del cielo cuando llueve), pero su composición química no es H_2O, es XYZ—a esta especie de "agua" la llamaremos 'bi-agua'.

Putnam argumenta que la bi-agua no es agua. Al fin y al cabo, si una expedición espacial trajera a la Tierra muestras de lo que en la Tierra Gemela llaman 'agua', cuando, pruebas químicas de por medio, descubriéramos que su composición química no es H_2O sino XYZ, concluiríamos que aquello que nos han traído desde ese planeta lejano no es agua. Ahora bien, si la bi-agua no es agua, entonces el término 'agua' tiene extensiones distintas cuando la profieren los habitantes de la Tierra y cuando la profieren los habitantes de la Tierra Gemela y, sobre la base de (II), si la extensión de 'agua' en boca de unos y de otros difiere, también lo hace así su intensión.

Sigamos describiendo el ejemplo: supongamos que $Oscar_1$ habita la Tierra, y $Oscar_2$ la Tierra Gemela. $Oscar_1$ y $Oscar_2$ son "idénticos"—entre otras cosas, ninguno de los dos sabe cuál es la composición química de la "sustancia acuosa" que lo rodea.

> Podemos suponer que no hay creencia que tenga $Oscar_1$ sobre el agua que no tenga $Oscar_2$ sobre la bi-agua. Si así se desea, podemos incluso suponer que $Oscar_1$ y

[13] Alguien podría protestar que el ejemplo, así expuesto, es imposible, porque a una especie le es esencial su cadena evolutiva: es imposible que dos cadenas causales o evolutivas distintas tengan como producto la misma especie. Por eso, los "humanos" de la Tierra Gemela no serían humanos, ni los "perros" perros—y esta crítica no se reduciría a las clases naturales o biológicas: por razones similares, el "castellano" que hablan los habitantes de la Tierra Gemela no es castellano, ni los "italianos" que la habitan italianos. Esto no supone un gran problema, ya que al ejemplo le bastaría que aquellos "humanos" y "perros" de la Tierra Gemela fueran lo suficientemente parecidos a humanos y perros. Por eso, cuando a lo largo de este trabajo entremos a describir o discutir ejemplos de Tierra Gemela pasaremos por alto estos detalles.

> $Oscar_2$ son duplicados exactos en apariencia, sensaciones, pensamientos, monólogo interior, etc.[14]

$Oscar_1$ y $Oscar_2$ están en el mismo estado psicológico (es esto algo que podemos *estipular* si, con Putnam, aceptamos que las propiedades internas determinan por completo las propiedades psicológicas (mentales)). Si están en el mismo estado psicológico, entre otras cosas, las intensiones que relacionan con los distintos términos serán exactamente las mismas (sobre la base de la asunción (I)); y estos términos tendrán también exactamente la misma extensión (sobre la base de (II)). Pero hemos dicho antes que los usos de 'agua' en la Tierra y en la Tierra Gemela difieren en extensión: hemos llegado a una contradicción.

Los ejemplos de Tierra Gemela demuestran que (I) y (II) no pueden sostenerse al mismo tiempo, al menos una de ellas habrá de ser falsa. Las propiedades internas de un sujeto no bastan para determinar cuáles son los contenidos y las extensiones de los términos que emplea:

> Ponlo como te dé la gana, 'los significados' simplemente no están en la *cabeza*![15]

Putnam (1975) concluye que los términos de clase natural como 'agua' son una especie de indéxicos; al igual que sucede con 'yo', 'aquí' o 'ahora', la extensión de 'agua' varía dependiendo de en qué contexto se profiere. Es por esta naturaleza indéxica que ambos principios (I) y (II) no pueden ser verdaderos sobre términos como 'agua': o bien negamos que el concepto que relaciona uno con 'agua' determina su contenido o, si no, que el contenido de 'agua' basta para determinar su extensión:

> De cualquier modo, la teoría de que los términos de clase natural como 'agua' son indéxicos deja la puerta abierta a decir que 'agua' en el dialecto del castellano de la Tierra Gemela tiene el mismo *significado* que 'agua' en el dialecto de la Tierra y una extensión diferente (que es lo que normalmente decimos sobre 'yo' en distintos idiolectos), abandonando así la doctrina de que 'el significado (la intensión) determina la extensión'; o decir, como hemos optado nosotros, que la diferencia en la extensión es *ipso facto* una diferencia en significado para los términos de clase natural, abandonando así la doctrina de que los significados son conceptos, o, es más, entidades mentales de *cualquier* clase.[16]

---

[14] You may suppose that there is no belief that $Oscar_1$ had about water that $Oscar_2$ did not have about 'water'. If you like, you may even suppose that $Oscar_1$ and $Oscar_2$ are exact duplicates in appearance, feelings, thoughts, interior monologue, etc. (Putnam (1975), p. 224)

[15] Cut the pie any way you like, 'meanings' just ain't in the *head*! (Putnam (1975), p. 227)

[16] The theory that natural-kind words like 'water' are indexical leaves it open, however, whether to say that 'water' in the Twin Earth dialect of English has the same *meaning* as 'water' in the Earth dialect and

Putnam asume así que el estado mental en el que se encuentra uno no basta para determinar el contenido del término en cuestión—su significado. $Oscar_1$ y $Oscar_2$ comparten sus propiedades internas y, de acuerdo con Putnam, también comparten conceptos. Pero 'agua' expresará contenidos distintos dependiendo de quién de los dos lo haya proferido—hé ahí la indexicalidad de los términos como 'agua'. Los significados no están en la cabeza.

Una exposición mínimamente extensa de las muchas tesis adoptadas por Putnam (1975) sobre (términos de) clases naturales no tiene cabida en esta breve sección. Enumeraremos algunas de las tesis que defiende, esperemos que esto sirva al menos como esbozo de la posición que sostiene:

- Diferenciemos entre: la descripción de las propiedades superficiales (observables) típicas de una clase, su *estereotipo* (en el caso del agua, algo así como "líquido transparente e insípido que bebemos y cae del cielo cuando llueve"), y la muestra que conforman algunas instancias paradigmáticas de la clase, su *estándar* (distintas muestras de agua que podemos señalar mediante ostensión).

- Tener las propiedades identificadas por el estereotipo de una clase no es una condición ni necesaria ni suficiente para caer dentro de la extensión de esa clase (ser líquido, transparente, etc. no es ni necesario ni suficiente para ser agua). Lo que determina que algo sea de una clase es estar en cierta relación de mismidad con las muestras que forman el estándar.

- Así, del mismo modo, la descripción que relaciona uno con un término de clase natural (el estereotipo) no basta para determinar la extensión del término. $Oscar_1$ y $Oscar_2$ dan las mismas condiciones para que algo sea "agua", pero el término 'agua' tiene diferentes extensiones dependiendo de quién de los dos lo profiera.

- Lo que determina la extensión de un término de clase natural proferido por un sujeto dado es generalmente cierta cadena causal-comunicativa en la que el sujeto toma parte.

---

a different extension (which is what we normally say about 'I' in different idiolects), thereby giving up the doctrine that 'meaning (intension) determines extension'; or to say, as we have chosen to do, that difference in extension is *ipso facto* a difference in meaning for natural-kind words, thereby giving up the doctrine that meanings are concepts, or, indeed, mental entities of *any* kind. (Putnam (1975), p. 234)

En los casos paradigmáticos, alguien *bautiza* unos ejemplares como formando el estándar de la clase. El término se pasa de un hablante a otro y el uso que hace un hablante del término refiere a lo mismo a lo que referían los usos que hacían del término aquéllos de los cuales lo recibió. El hablante no tiene por qué manejar una descripción de la clase que fije la extensión del término, es su participación en la cadena causal-comunicativa lo que lo hace.

- La relación de mismidad con las instancias que fijan el estándar se da entre distintos mundos posibles: una sustancia en cualquier mundo posible es agua si y sólo si guarda la relación de mismidad con las muestras del estándar, no tiene por qué tener las propiedades que fija el estereotipo; los términos de clase son *designadores rígidos*.

- Más allá de las propiedades superficiales comunes, hay cierta "estructura oculta" que comparten todas las instancias que caen bajo la extensión de una clase natural; esa estructura oculta le es esencial a la clase, aunque normalmente sólo la conocemos mediante investigación empírica (científica). Por ejemplo, al agua le es esencial ser H_2O (es falso que el agua sea H_2O en la Tierra y XYZ en la Tierra Gemela, algo es agua si y sólo si es H_2O). De todo esto se sigue que el enunciado 'El agua es H_2O' expresa una proposición necesaria, aunque sólo cognoscible *a posteriori*.

- A pesar de que lo que decide si algo es una instancia de cierta clase natural es que tenga esa estructura que le es esencial a la clase, el estereotipo juega un papel relevante para que varios sujetos puedan conocer si algo es una instancia de la clase. Ser líquido y transparente no es ni necesario ni suficiente para ser agua, pero marca un criterio fiable para saber si algo es agua o no. Por eso, no todo hablante competente ha de conocer qué es lo que en realidad determina que algo caiga en la extensión de la clase (es más, puede que de hecho *nadie* en la comunidad lingüística sepa esto). Lo que sí es necesario, para que uno use un término de clase competente y exitosamente, es que en su comunidad lingüística haya individuos que puedan determinar, fiablemente, si algo cae bajo la extensión de ese término o no. Esto es, hay lo que Putnam llama *división de trabajo en la comunidad lingüística*.

- Por eso, uno puede ser hablante competente del término de clase `haya` sin que sepa qué determina realmente que algo caiga bajo su extensión, o sin que sea capaz de

distinguir entre hayas y olmos—relacionará el mismo concepto con los dos términos. Ahora, cuando profiera enunciados que contengan el término 'haya', éstos serán sobre hayas (no sobre olmos), gracias a que los usos que hace el hablante descansan sobre la división de trabajo que se realiza en su comunidad lingüística. Uno no ha de manejar una descripción que determina correctamente la extensión del término 'c' para poder ser considerado competente en cuanto a ese término.

- Esto abre la puerta a que uno tenga en su vocabulario términos que, inadvertidamente para él, son sinónimos. Un hablante competente de castellano e inglés puede tener en su vocabulario los términos 'haya' y 'beech', pero no saber que algo es un haya si y sólo si es un *beech*.

Esto es, resumiendo:

> Hemos visto que la extensión de un término no se fija mediante algún concepto que el hablante individual tiene en su cabeza, y esto es verdad porque la extensión se determina en general *socialmente* – hay una división del trabajo lingüístico tanto como trabajo 'real' – y porque la extensión se determina en parte *indéxicamente*. La extensión de nuestros términos depende la naturaleza actual de los objetos particulares que sirven como paradigma, y esta naturaleza actual no es, en general, totalmente conocida por el hablante. La teoría semántica tradicional deja fuera sólo dos contribuciones a la determinación de la extensión – la contribución de la sociedad y la contribución del mundo real![17]

Terminemos. Putnam (1975) es externista sobre qué determina cuál es el contenido expresado mediante la proferencia de un término de clase natural pero, como ya advertíamos al comienzo de esta sección, no lo es en cuanto a qué determina cuál es el contenido de un estado mental. Así, por ejemplo, según él $Oscar_1$ y $Oscar_2$ estarán en el mismo estado mental (con los mismos conceptos, creencias y pensamientos) en cuanto están en el mismo estado interno. Esta idea le lleva a mantener que las hipotéticas proferencias de $Oscar_1$ y $Oscar_2$ de 'El agua es insípida' dan a conocer una misma creencia que comparten, aunque las proferencias mismas difieren en contenido y extensión. Esta consecuencia es extraña, se sigue que $Oscar_1$ y $Oscar_2$ tienen la misma creencia, pero que la del primero es sobre el agua y la del segundo sobre la bi-agua.

[17] We have now seen that the extension of a term is not fixed by a concept that the individual speaker has in his head, and this is true both because extension is, in general, determined *socially* – there is division of linguistic labor as much as of 'real' labor – and because extension is, in part, determined *indexically*. The extension of our terms depends upon the actual nature of the particular things that serve as paradigms, and this actual nature is not, in general, fully known to the speaker. Traditional semantic theory leaves out only two contributions to the determination of extension – the contribution of society and the contribution of the real world! (Putnam (1975), p. 245)

Bien, Putnam no defiende un modelo externista para nuestros estados mentales (es extraño que no use el mismo experimento mental de Tierra Gemela con ese fin), pero es *éste* el modelo en el que estamos interesados nosotros. Este modelo diría, más o menos, que qué clases contiene el entorno de un sujeto es relevante para determinar qué conceptos tiene, cuáles son los contenidos de sus estados mentales intencionales. Así, según el externismo de clases, dos sujetos pueden estar en el mismo estado interno, pero tener conceptos o creencias distintas porque sus respectivos entornos contienen clases naturales distintas.

$Oscar_1$ y $Oscar_2$ estarán en distintos estados mentales porque hay dos clases distintas (el agua y la bi-agua) que habitan sus respectivos entornos. El primero tiene el concepto AGUA, el segundo el concepto BI-AGUA, los cuales difieren en extensión (aunque $Oscar_1$ y $Oscar_2$ relacionan el mismo estereotipo con ellos); $Oscar_1$ tendrá la creencia de que el agua moja y el temor de ahogarse en una piscina de agua; $Oscar_2$ la creencia de que la bi-agua moja y el temor de ahogarse en una piscina de bi-agua. Da igual que se encuentren en la misma situación interna, fenoménica o física: esas propiedades no bastan para determinar sus estados mentales.

Por lo demás, el externista de clases naturales que estamos esbozando podrá adherirse a la mayoría de las afirmaciones hechas por Putnam que hemos presentado más arriba, y podrá proponer su equivalente para los conceptos de clases naturales. Por ejemplo, el externista de clases defenderá que uno no ha de conocer qué diferencia un haya de un olmo para poder tener el concepto HAYA, y que la adquisición de este concepto (así como del término 'haya') descansa en la división del trabajo lingüístico, la presencia de hayas en nuestro entorno, o la participación en cierta cadena causal-comunicativa.

Mencionemos al menos que hay una tesis que defendió Putnam (1975) que seguramente rechazan la mayoría de los externistas: es falso que los términos de clase natural sean algún tipo de indéxicos. Hay un motivo importante para negar esta afirmación de Putnam, y es que, a diferencia de lo que pasa con términos indéxicos paradigmáticos como 'yo', 'aquí' y 'ahora', los términos de clase natural no tienen una regla que determina la referencia del término en cada proferencia y que ha de conocer el hablante para poder ser considerado competente. Por eso, la mayoría de externistas diría que $Oscar_1$ y $Oscar_2$ no usan el mismo término 'agua' (pero que los distintos usos de este

mismo término tienen referencias distintas), sino que emplean dos términos distintos (con contenidos y referencias distintas) que resultan ser homofónicos.

Las descripciones que relaciona uno con sus conceptos de clase natural no determinan a qué refieren esos conceptos—como proponía el modelo tradicional internista que hemos presentado antes). Las propiedades internas de uno no bastan para determinar qué conceptos (qué creencias) tiene. Esto depende en parte de cuáles son las clases naturales presentes en su entorno, o de cuáles son las instancias que forman el estándar del término que relaciona con ese concepto.

Pasemos a caracterizar el externismo social de Tyler Burge.

0.2. TYLER BURGE: EXTERNISMO SOCIAL, COMPRENSIÓN INCOMPLETA Y ADQUISICIÓN DEFERENCIAL

Tyler Burge escribió "Individualism and the Mental" en 1979, (podría decirse que) inaugurando la corriente de pensamiento que nosotros llamamos 'externismo semántico'. Influenciado en gran medida por textos de Kripke y, sobre todo, Putnam, el artículo da un paso más allá en el rechazo del modelo internista que hemos caracterizado páginas más arriba.

La parte más importante del artículo está formada por un experimento mental en tres pasos, al estilo de las tierras gemelas de Putnam; el argumento basado en el experimento concluye que el internismo semántico es falso. El primer paso del experimento nos presenta a un individuo que guarda un número amplio de actitudes que le adscribiríamos usando cláusulas de contenido que contienen el término 'artritis', que ocurre *oblicuamente* (algunos párrafos más abajo explicaremos qué quiere decir que un término ocurra oblicuamente en una cláusula de contenido). Por ejemplo cree: que la artritis es dolorosa, pero que la artritis es menos grave que el cáncer, o que al igual que su padre, él ha empezado a padecer artritis a los cincuenta y ocho años de edad; todas ellas son creencias de hecho verdaderas.

Pero el individuo también tiene una creencia falsa: que padece artritis en el muslo. Esta creencia es falsa *por definición*, ya que la artritis es, por definición, inflamación de las articulaciones; uno sólo puede tener artritis en las articulaciones (por definición del término 'artritis'). El sujeto que estamos describiendo acude al médico y le explica que teme estar padeciendo artritis en el muslo. Por supuesto, el médico le responde que uno sólo puede tener artritis en las articulaciones (cualquier diccionario podría haberle enseñado esto), y que la suya será seguramente otra enfermedad reumática. El paciente corrige su actitud, abandona su creencia de que tiene artritis en el muslo, y comienza a preocuparse por cuál será la naturaleza de los dolores que siente en el muslo.

El segundo paso del experimento nos propone un escenario contrafáctico.

> Concibamos una situación en la cual el paciente procede, desde su nacimiento, por el mismo curso de eventos físicos que de hecho vive actualmente, justo hasta incluir el momento en que por primera vez da a conocer sus temores al médico. Le suceden justo las mismas cosas (descritas no-intencionalmente). Tiene la misma historia fisiológica, las mismas enfermedades, las mismas vivencias físicas internas. Realiza los mismos movimientos, adopta la misma conducta, recibe los mismos estímulos sensoriales (descritos fisiológicamente). (...) Dice y oye las mismas palabras (formas de palabras) en los mismos momentos en que las dice y oye actualmente. Desarrolla la disposición a asentir a 'La artritis puede desarrollarse en el muslo', y 'Tengo artritis en el muslo' como resultado de las mismas causas descritas físicamente. Imaginemos que la experiencia fenoménica no-intencional del paciente es la misma. Tiene los mismos dolores, campos visuales, imágenes y articulaciones verbales internas. La *contrafacticidad* en la suposición afecta sólo al entorno social del paciente. En el escenario actual, 'artritis', como se usa en su comunidad, no se aplica a achaques fuera de las articulaciones. Es más, no lo hace así dada cierta definición estándar, no-técnica. Pero en nuestro caso imaginario, los médicos, lexicógrafos y gente corriente pero informada aplican 'artritis' no sólo a la artritis, sino a otros varios achaques reumáticos.[18]

[18] We are to conceive of a situation in which the patient proceeds from birth through the same course of physical events that he actually does, right to and including the time at which he first reports his fear to his doctor. Precisely the same things (non-intentionally described) happen to him. He has the same physiological history, the same diseases, the same internal physical occurrences. He goes through the same motions, engages in the same behavior, has the same sensory intake (physiologically described). (...) He says and hears the same words (word forms) at the same times he actually does. He develops the disposition to assent to 'Arthritis can occur in the thigh' and 'I have arthritis in the thigh' as a result of the same physically described proximate causes. (...) We further imagine that the patient's non-intentional, phenomenal experience is the same. He has the same pains, visual fields, images, and internal verbal rehearsals. The *counterfactuality* in the supposition touches only the patient's social environment. In actual fact, 'arthritis', as used in his community, does not apply to ailments outside joints. Indeed, it fails to do so by a standard, non-technical dictionary definition. But in our imagined case, physicians, lexicographers, and informed laymen apply 'arthritis' not only to arthritis but to various other rheumatoid ailments. (Burge (1979), p. 105)

Dicho de forma menos detallada: el segundo paso del experimento viene a decir que el individuo podría haber tenido la misma historia física y los mismos fenómenos mentales no-intencionales, mientras que el término 'artritis' se hubiera definido de modo distinto en su comunidad lingüística, tal que se aplicara también a dolores fuera de las articulaciones. Incluso en ese caso el individuo tendría las mismas disposiciones a proferir y asentir a los mismos enunciados ("formas" de enunciados).

El último paso del experimento es una interpretación del caso contrafáctico descrito:

> En la situación contrafáctica, el paciente carece de algunas—probablemente *todas*—de las actitudes comúnmente atribuidas con cláusulas de contenido que contienen 'artritis' ocurriendo oblicuamente. (...) Resulta difícil ver cómo el paciente podría haber adquirido la noción de artritis. El término 'artritis' no significa *artritis* en la comunidad contrafáctica. No se aplica sólo a inflamaciones en las articulaciones. (...) Nuestras adscripciones de cláusulas de contenido al paciente (y las adscripciones en su comunidad) no constituirían atribuciones de los mismos contenidos que atribuimos en el escenario actual.(...) Por lo tanto, las actitudes de contenido contrafácticas del paciente difieren de las que tiene actualmente.[19]

Si aceptamos los tres pasos del experimento (que en la situación actual el paciente tiene creencias en parte constituidas por el concepto ARTRITIS; que podría haber tenido la misma historia psico-física descrita no-intencionalmente aun y cuando 'artritis' se describiera de un modo diferente; y que en un escenario tal no tendría actitudes proposicionales que pudiéramos describir usando adscripciones que contienen apariciones oblicuas de 'artritis'), se sigue que el internismo es falso. El ejemplo muestra que uno podría estar en la misma situación psico-física (descrita no-intencionalmente), en la misma situación interna, habiendo vivido las mismas vivencias, pero tener actitudes de contenido distintas si su entorno social fuera distinto, si habitara una comunidad lingüística distinta. Esto es, los estados internos del individuo (sus propiedades tanto físicas como fenoménicas) no bastarían para determinar qué piensa, cree, teme o desea.

[19] In the counterfactual situation, the patient lacks some—probably *all*—of the attitudes commonly attributed with content clauses containing 'arthritis' in oblique occurrence. (...) It is hard to see how the patient could have picked up the notion of arthritis. The word 'arthritis' in the counterfactual community does not mean *arthritis*. It does not apply only to inflammations of joints. (...) Our ascriptions of content clauses to the patient (and ascriptions within his community) would not constitute attributions of the same contents we actually attribute. (...) So the patient's counterfactual attitude contents differ from his actual ones. (Burge (1979), p. 106)

Al comienzo de este capítulo hemos bosquejado cierta posición, que sostiene algunas tesis concretas sobre determinación del contenido y de la referencia. Este modelo será erróneo si la caracterización de Burge de su experimento mental es correcta:

> Los ejemplos de Tierra Gemela (como los ejemplos de "Individualism and the Mental") indican que el orden de explicación no sigue una línea recta desde las actitudes proposicionales "en sentido estrecho" hasta las extensiones de los términos. Por contra, para conocer y explicar qué cree una persona *de dicto*, uno debe típicamente saber algo sobre qué cree *de re*, sobre qué creen *de re* (y *de dicto*) sus coetáneos, sobre a qué entidades hacen ostensión, sobre qué significan las palabras que usan él y sus coetáneos, y sobre qué entidades caen dentro de las extensiones de sus términos.[20]

Por ahora dejaremos de lado las creencias *de re,* las ostensiones y el rol de la referencia a la hora de determinar los contenidos y pensamientos de alguien—nos bastará con el rol que juegan las relaciones lingüístico-sociales.

Burge no se limita a concluir que el internismo es falso. También sugiere qué tipo de entidades o relaciones pueden ser relevantes a la hora de determinar en qué estado mental intencional se encuentra un individuo: qué contexto social o lingüístico habita, por ejemplo, es relevante para determinar cuáles son los contenidos de sus estados mentales intencionales.

> El resultado de estas disquisiciones es que los contenidos mentales del paciente difieren, mientras que todas sus historias físicas y mentales no-intencionales, consideradas aisladas de su contexto social, se mantienen. (...) Las diferencias parecen surgir de diferencias que se encuentran "fuera" del paciente considerado como un aislado organismo físico, mecanismo causal o aposento de la consciencia. La diferencia en sus contenidos mentales es atribuible a diferencias en su entorno social.[21]

La tesis defendida viene a ser que en qué estado mental intencional se encuentra uno, qué piensa o qué conceptos tiene, en parte depende de cómo es su entorno social y lingüístico; "tenemos que tener en cuenta la comunidad de una persona al interpretar sus

[20] The twin-earth examples (like the examples from 'Individualism and the Mental') indicate that the order of explication does not run in a straight line from propositional attitudes in the 'narrow sense' to the extensions of terms. Rather, to know and explicate what a person believes *de dicto*, one must typically know something about what he believes *de re*, about what his fellows believe *de re* (and *de dicto*), about what entities they ostend, about what he and his fellows' words mean, and about what entities fall in the extensions of their terms. (Burge (1982), p. 95)

[21] The upshot of these reflections is that the patient's mental contents differ, while his entire physical and non-intentional mental histories, considered in isolation from their social context, remain the same. (...) The differences seem to stem from differences 'outside' the patient considered as an isolated physical organism, causal mechanism, or seat of consciousness. The difference in his mental contents is attributable to differences in his social environment. (Burge (1979), p. 106)

palabras y describir sus actitudes"[22]. En principio dos individuos pueden ser idénticos *internamente*, pero diferir en sus estados mentales intencionales, por el mero hecho de que interactúan con dos comunidades sociales (lingüísticas) distintas. A esta tesis la llamaremos *externismo social*.

Llamemos la atención sobre una consecuencia del externismo social, una presuposición del experimento tal y como lo hemos presentado (y lo presentó Burge): uno puede comprender *parcialmente* un concepto que tiene (o un término que usa de forma competente).

Es necesario distinguir entre dos usos de 'comprender un término'. Según el primero, uno ya comprende un término si lo hace de un modo mínimo, si el uso que hace de él es suficiente para que le adscribamos pensamientos y creencias usando el término oblicuamente, si su comprensión es suficiente para que le adscribamos el concepto en cuestión. Según el segundo, uno comprende un término o concepto sólo si conoce la definición que delimita qué entidades caen (y cuáles no) dentro de la extensión del término o concepto. Y uno puede tener el primer tipo de comprensión sin tener el segundo: uno puede tener un concepto sin ser capaz de articular la norma que delimita qué cae dentro del concepto y qué no, sin que conozca la definición exacta del concepto:

> Uno puede pensar con un concepto incluso cuando uno no lo ha dominado completamente, en el sentido de que uno le asocia una concepción (o explicación conceptual) incorrecta.[23]

Es el caso del individuo que hemos descrito en el primer paso del experimento. Éste no sabe que, por definición, uno puede tener artritis sólo en las articulaciones, no sabe qué es lo que hace que una dolencia concreta sea artritis. Pero al exponer su caso no hemos dudado en adscribirle el concepto ARTRITIS. Alguien podría negar que uno puede tener sólo una comprensión parcial de sus conceptos, podría negar que el paciente que hemos descrito posee el concepto ARTRITIS (probablemente, muchos internistas lo harían). Si nuestras propiedades internas determinaran qué conceptos tenemos, resulta plausible

[22] We have to take account of a person's community in interpreting his words and describing his attitudes. (Burge (1979), p. 113)
[23] One may think with a concept even though one has incompletely mastered it, in the sense that one associates a mistaken conception (or conceptual explication) with it. (Burge (1993b), p. 298)

que también las explicaciones que diéramos sobre esos conceptos determinaran qué conceptos tenemos. El externismo social niega este último punto.

> Cuando la comprensión de un sujeto es todavía parcial, algunas veces atribuimos contenidos mentales según los mismos términos que el sujeto todavía tiene que dominar. Las posiciones tradicionales entienden que dominar un término consiste en unirlo con un concepto ya dominado (o innato). Pero parecería, en cambio, que muchos conceptos (o componentes de contenidos mentales) son como palabras, que se emplean antes de que se dominan. En ambos casos, el empleo parece ser una parte integral del proceso de dominio.[24]

Y creemos preciso subrayar que ésta es una diferencia importante que guarda el externismo social con el internismo. En el modelo internista la explicación conceptual que puede dar el sujeto define cuál es el concepto que tiene y determina la referencia de ese concepto. En el modelo externista social los usos lingüísticos que hacen los coetáneos lingüísticos del sujeto son relevantes para determinar qué conceptos tiene éste. Esta tesis asume que es posible que uno no sea capaz de dar la explicación más correcta sobre la naturaleza de los conceptos que tiene; es en este sentido que un sujeto puede tener una comprensión *parcial* de sus conceptos. La adquisición de un concepto u otro por parte del sujeto descansa en parte en sus relaciones lingüístico-sociales, y puede adquirir un concepto *deferencialmente*.

> Cuando el individuo defiere a otros, no lo hace así siempre para ajustar o fijar la referencia, sino para ajustar el conocimiento explicativo que tiene el individuo acerca de la referencia que ya está fijada. (...) En algunos casos, las habilidades explicatorias de un individuo no sólo no bastan para fijar la referencia del término del individuo; no agotan el significado expresado por el término en el idiolecto del individuo.[25]

Otra idea que presupone el experimento es que las adscripciones de actitudes que contienen términos que aparecen *oblicuamente* describen de cierta manera especial el estado mental intencional en el que se encuentra el individuo.

---

[24] While the subject's understanding is still partial, we sometimes attribute mental contents in the very terms the subject has yet to master. Traditional views take mastering a word to consist in matching it with an already mastered (or innate) concept. But it would seem, rather, that many concepts (or mental content components) are like words, in that they may be employed before they are mastered. In both cases, employment appears to be an integral part of the process of mastery. (Burge (1979), p. 107, nota a pie de página 1)

[25] When the individual defers to others, it is not in all cases to sharpen or fix the reference, but to sharpen the individual's explicative knowledge of a referent that is already fixed. (...) In some cases, an individual's explicational ability not only does not suffice to fix the referent of the individual's word; it does not exhaust the meaning expressed by a word in the individual's idiolect. (Burge (1989), p. 282)

Diferenciemos entre apariciones *transparentes* y apariciones *oblicuas* de los términos en las adscripciones de actitudes proposicionales. La aparición de un término en un contexto tal será transparente si y sólo si el término es sustituible por cualquier otro término con la misma extensión *salva veritate*. Si no, la aparición será oblicua. Pongamos un ejemplo. Antonio, torpe y desaliñado, escribe anónimas cartas de amor a Jenny, la joven oficinista de su trabajo. Conmovida por el sentimentalismo de las cartas, Jenny opina que quien las haya escrito (sea quien sea), ha de ser un experto amante. Consideremos el siguiente enunciado:

(Jenny) Jenny opina que el autor de las cartas es un amante experto.

(Jenny) es verdadero y, como Antonio es el autor de las cartas que recibe Jenny, 'Antonio' y 'El autor de las cartas' tienen la misma extensión, refieren exactamente al mismo objeto, a Antonio. La cuestión es que (Jenny) tiene relacionados dos usos distintos. El primero es tal que, si sustituyéramos 'El autor de las cartas' por 'Antonio', lo dicho seguiría siendo verdadero (ya que Jenny opina acerca de quien haya escrito las cartas, de hecho Antonio, que es un amante experto). En este primer uso, la aparición de 'El autor de las cartas' es, pues, *transparente*. Por contra, el segundo uso de (Jenny) es tal que, si sustituyéramos 'El autor de las cartas' por 'Antonio', lo dicho sería falso (ya que, dado su aspecto y su tartamudeo cada vez que se dirige a ella, Jenny no opina que Antonio sea un gran amante, si le preguntáramos 'Jenny, ¿crees que Antonio es un amante experto?' respondería '¡¡No!!'). La aparición de 'El autor de las cartas' en este segundo uso es *oblicua*.

La cuestión es que, según Burge, las adscripciones de actitudes proposicionales que contienen la aparición oblicua de algún término, cumplen con un cometido especial:

> Los términos que ocurren oblicuamente en las cláusulas de contenido son el modo principal para identificar los estados o eventos mentales intencionales de una persona. Es aconsejable remarcar un detalle aquí. Es normal suponer que aquellas cláusulas de contenido que se pueden adscribir correctamente a una persona y que no son en general intersustituibles *salva veritate*—y, ciertamente, aquellas que contienen expresiones-contraparte extensionalmente no-equivalentes—identifican diferentes estados o eventos mentales.[26]

[26] ...obliquely occurring expressions in content clauses are a primary means of identifying a person's intentional mental states or events. A further point is worth remarking here. It is normal to suppose that those content clauses correctly ascribable to a person that are not in general intersubstitutable *salva*

> Es más, parece inobjetable defender que las expresiones que ocurren oblicuamente en las atribuciones de actitudes proposicionales son extremadamente importantes para caracterizar el estado mental de una persona. Estas apariciones son la materia con la cual están hechas las explicaciones de sus acciones y las aseveraciones de su racionalidad.[27]
>
> El objetivo de estas atribuciones es caracterizar los eventos y estados mentales de un sujeto de tal modo que se explique el *modo* en que opina o piensa sobre los objetos en su entorno.[28]

Cuando juzgamos que el paciente cree que la artritis es dolorosa, o cuando juzgamos que el paciente contrafáctico no cree que la artritis es dolorosa, el término 'artritis' aparece oblicuamente en nuestras adscripciones. Si Burge está en lo cierto, pues, estas adscripciones describirán en qué estado mental se encuentran los dos protagonistas del experimento, el *modo* en que entienden su entorno. Si una adscripción es verdadera de uno pero falsa del otro, se seguirá, por lo tanto, que estos dos individuos se encuentran en estados mentales diferentes, que entienden sus respectivos entornos según dos *modos* distintos, que guardan perspectivas distintas.

Sigamos. El argumento se basa en un experimento que presenta a dos individuos en situaciones concretas; cabe preguntarse hasta qué punto se extiende la tesis externista. Por ejemplo, uno podría sospechar que la tesis sólo se aplica en los casos que hay una comprensión *parcial* de un concepto. Estaría equivocado. El experimento también demuestra que las propiedades internas de uno no bastan para determinar si tiene una comprensión parcial o completa del concepto que expresa al proferir el término 'artritis'; necesitamos comparar esa comprensión con el uso del término que se hace en su comunidad lingüística para determinar qué nivel de comprensión tiene. Si necesitamos de factores externos para determinar que la comprensión que tiene S de su concepto C es completa, no podemos decir que cuando uno tiene una comprensión completa de C podemos determinar que tiene ese concepto sin acudir a factores externos. Así, vemos que "*incluso esas actitudes proposicionales no infectadas por la*

---

veritate—and certainly those that involve extensionally non-equivalent counterpart expressions—identify different mental states or events. (Burge (1979), p. 104)

[27] Moreover, it seems unexceptionable to claim that the obliquely occurring expressions in propositional attitude attributions are critical for characterizing a given person's mental state. Such occurrences are the stuff of which explanations of his actions and assessments of his rationality are made. (Burge (1982), p. 84)

[28] The point of such attributions is to characterize a subject's mental states and events in such a way as to take into account the *way* he views or thinks about objects in his environment. (Burge (1982), p. 92)

comprensión incompleta dependen para su contenido de factores sociales que son independientes del individuo, descrito asocial y no-intencionalmente"[29].

La tesis externista en su vertiente social, pues, es verdadera acerca de todo tipo de concepto que admita de una comprensión parcial, y para cualquier pensamiento o creencia en parte constituido por un concepto tal. Para cualquier concepto de este tipo, su naturaleza depende de qué relaciones guardamos con qué comunidades sociales y lingüísticas. La gran mayoría de nuestros conceptos y pensamientos depende para su individuación de las relaciones lingüístico-sociales que guardamos.

Nos gustaría terminar esta sección con una advertencia que hace Burge. Es común entender el externismo como una teoría sobre la naturaleza del contenido—esta interpretación es errónea. El externismo semántico es una tesis acerca de la naturaleza de nuestros estados mentales y conceptos:

> El anti-individualismo no trata fundamentalmente sobre la naturaleza del contenido. Trata sobre la naturaleza de los estados y eventos mentales representacionales. Trata sobre las condiciones esenciales o constitutivas para que un individuo tenga el tipo de estados y eventos mentales que tiene. (...) La conclusión trata sobre los pensamientos mismos. Trata sobre cómo *tener* ciertos pensamientos depende constitutivamente de relaciones con el entorno. No es sobre la naturaleza de los contenidos mentales mismos.[30]

El externismo no dice nada acerca de qué tipo de entidad son los contenidos (ni los lingüísticos ni los mentales), no es una tesis sobre su naturaleza. Lo que sí dice el externismo es qué factores condicionan qué contenido tiene un pensamiento dado, y qué contenidos no. En cuanto los pensamientos y conceptos dependen constitutivamente de sus contenidos, el externismo es una tesis sobre la naturaleza de nuestros pensamientos y conceptos.

Terminemos esta sección introductoria enumerando algunas de las tesis que defiende Burge, las cuales creemos que caracterizan adecuadamente al externista social:

---

[29] *...even those propositional attitudes not infected by incomplete understanding* depend for their content on social factors that are independent of the individual, asocially and non-intentionally described. (Burge (1979), pp. 112-113)

[30] Anti-individualism is not fundamentally about the nature of content. It is about the nature of representational mental states and events. It is about constitutive or essential conditions on an individual's having the kinds of mental states and events that the individual has. (...) The conclusion is about the thoughts themselves. It is about how *having* certain thoughts constitutively depends on relations to the environment. It is not about the nature of the thought contents themselves. (Burge (2007b), pp. 155-156)

- Las propiedades fenoménico-físicas de un sujeto, su conducta, sus disposiciones (todas ellas individuadas no-intencionalmente), es decir: sus propiedades internas no bastan para determinar qué conceptos y pensamientos tiene, para determinar qué contenidos tienen sus estados mentales intencionales. El internismo es falso.
- Entre los factores relevantes para individuar los conceptos y pensamientos de un sujeto están las relaciones lingüístico-sociales que guarda. Nuestros pensamientos y conceptos dependen constitutivamente de las comunidades lingüísticas que habitamos.
- Una cosa es qué concepto tiene un sujeto y otra qué explicación daría el sujeto de ese concepto (lo segundo no basta para determinar lo primero). Uno no tiene por qué conocer las definiciones de los términos que emplea para hacer un uso competente de esos términos. Es común que uno tenga una comprensión parcial de sus propios conceptos.
- Adquirimos varios de nuestros conceptos mediante cadenas comunicativas, *deferencialmente*.

Resumiendo,

> Ningún fenómeno mental intencional de ningún hombre es insular. Todo hombre es una pieza del continente social, una parte de la tierra firme social.[31]

0.3. EXTERNISMO SINGULAR Y NEO-FREGEANO

Comencemos esta sección esbozando cómo explica, el modelo tradicional, que alguien tenga un pensamiento sobre un objeto particular. Según este modelo, el único modo que tiene uno de pensar acerca de un objeto es *descriptivamente*, ejemplificando un *pensamiento descriptivo*. Un pensamiento descriptivo es un pensamiento del tipo EL (ÚNICO) ϕ ES F, donde la descripción definida EL ϕ determina cuál es el objeto sobre el que trata el pensamiento: será sobre aquel objeto que satisfaga la descripción (cualquiera que sea este objeto). Si, de hecho, *x* es el único objeto que satisface la descripción ϕ, entonces el pensamiento será sobre *x* (ahora, si en un escenario

[31] No man's intentional mental phenomena are insular. Every man is a piece of the social continent, a part of the social main. (Burge (1979), p. 116)

contrafáctico no es *x*, sino *y*, el que satisface ϕ, entonces en ese escenario el mismo pensamiento EL (ÚNICO) ϕ ES F será sobre *y*, no sobre *x*).

Por ejemplo, hay una descripción que el astrónomo babilonio Hammurabi relaciona con Héspero, a saber, que es el primer cuerpo celeste visible al atardecer. Por lo tanto, cuando decimos que Hammurabi está pensando que Héspero es una estrella, es el pensamiento descriptivo EL PRIMER CUERPO CELESTE VISIBLE AL ATARDECER ES UNA ESTRELLA el que está ejemplificando; dado que la descripción "El primer cuerpo celeste visible al atardecer" de hecho refiere a Héspero, su pensamiento será sobre ese mismo objeto.

El externista singular afirma que, además de descriptivamente, uno puede pensar acerca de un objeto ejemplificando un *pensamiento singular*. Seguramente, la discusión entre externistas singulares y los defensores del modelo internista se puede plantear como una discusión acerca de si hay o no pensamientos singulares, y a lo largo de esta sección nos dedicaremos en gran parte a caracterizar qué clase de pensamientos son estos supuestos pensamientos singulares. Por de pronto, podemos decir que a los pensamientos singulares les es esencial el objeto al cual refieren. La referencia es relevante a la hora de determinar cuál es el pensamiento singular que tiene un sujeto, es imposible tener el mismo pensamiento singular sobre objetos distintos. Así, los pensamientos descriptivos y los singulares difieren en propiedades modales. El pensamiento EL (ÚNICO) ϕ ES F de S será sobre *x* en este mundo posible, pero puede ser sobre *y* en otro, y sobre *z* en otro. En cambio, el pensamiento singular X ES F será sobre *x* en todo mundo posible.

Supongamos, por ejemplo, que Hammurabi tiene un *doppelgänger* en la Tierra Gemela, Bammurabi. El primer cuerpo celeste visible en los atardeceres de la Tierra Gemela ocupa, en el cielo de ese planeta, la misma posición que ocupa Héspero en el cielo de la Tierra, y sus habitantes usan el término 'Héspero' para referirse a él. Pero sucede que aquel objeto que los habitantes de la Tierra Gemela llaman 'Héspero' no es Héspero (Venus), sino otro planeta u otra estrella en (supongamos) una galaxia distinta a la Vía Láctea. Hammurabi y Bammurabi son internamente indistinguibles; los dos relacionan la misma descripción con el nombre propio tipo 'Héspero', "El primer cuerpo celeste visible por la tarde", y los dos tienen una creencia que expresarían profiriendo el enunciado tipo 'Héspero es una estrella'. Si los pensamientos que expresan mediante

esas proferencias fueran descriptivos, entonces Hammurabi y Bammurabi compartirían pensamiento (a saber: EL PRIMER CUERPO CELESTE VISIBLE AL ATARDECER ES UNA ESTRELLA). Pero, según el externista singular, los pensamientos que expresan Hammurabi y Bammurabi mediante sus respectivas proferencias son dos pensamientos singulares distintos, en el caso de Hammurabi HÉSPERO ES UNA ESTRELLA, en el caso de Bammurabi BI-HÉSPERO ES UNA ESTRELLA.

Así, el externista singular concluye que qué objetos o individuos contiene el entorno de un sujeto es relevante para determinar qué conceptos tiene, cuáles son los contenidos de sus estados mentales intencionales. Dos sujetos pueden estar en el mismo estado interno pero tener conceptos o creencias distintas porque sus respectivos entornos contienen objetos particulares distintos.

Es una cuestión a debatir qué condiciones ha de cumplir un sujeto (si alguna) para poder tener un pensamiento singular sobre un objeto *x*—sea cierta capacidad discriminatoria u otra condición epistémica, alguna relación de *acquaintance*, o simplemente tener a mano una descripción que de hecho refiere al objeto. La discusión es interesante e importante, pero no entraremos aquí a estas cuestiones[32]. Pero sí nos gustaría al menos esbozar dos modos de desarrollar esta idea de que hay pensamientos singulares, guiados respectivamente por un modelo neo-russelliano y otro neo-fregeano del contenido y la determinación de la referencia. Además, la caracterización de un modelo neo-fregeano de externismo singular (nos basaremos en Evans (1982) para ello) nos permitirá explicar cómo es posible defender un modelo externista pero de carácter fregeano.

La corriente neo-russelliana afirma que hay proposiciones singulares, esto es, proposiciones en parte constituidas por los objetos particulares acerca de los cuales son:

> Este análisis de la proposición expresada por 'John es alto' lo provee con dos componentes: la propiedad expresada por el predicado es alto, y el particular John. Así es, John mismo, ahí, atrapado en una proposición.[33]

[32] Véanse: Kaplan (1970, 1989), Burge (1977), Evans (1982) o Jeshion (2010).
[33] [This] analysis of the proposition expressed by 'John is tall' provides it with two components: the property expressed by the predicate is tall, and the individual John. That's right, John himself, right there, trapped in a proposition. (Kaplan (1970), p. 344)

Según el modelo neo-russelliano, la única aportación de un nombre propio a la proposición expresada por el enunciado en el que toma parte es (siempre) el objeto al que refiere. Por eso, el objeto mismo es parte constituyente de la proposición expresada, y no así el modo en que el sujeto (o el hablante) piensa acerca del objeto.

Por eso, si el nombre propio *a* y el nombre propio *b* refieren a un mismo objeto, se sigue que los enunciados *Fa* y *Fb* expresarán exactamente la misma proposición singular. Por ejemplo, dado que 'Héspero' y 'Fósforo' refieren al mismo objeto, a Venus, el enunciado 'Héspero es una estrella' y el enunciado 'Fósforo es una estrella' expresarán la misma proposición singular. Pero sucede que uno puede asentir al enunciado 'Héspero es una estrella' pero no así a 'Fósforo es una estrella' porque, por ejemplo, uno cree (falsamente) que Héspero no es Fósforo, que Héspero es una estrella, pero que Fósforo no lo es (es un planeta). Dado que, según el neo-russelliano, esos enunciados expresan la misma proposición, ha de explicar cómo es posible que uno asienta y no asienta al mismo tiempo a la misma proposición singular sin por ello caer en contradicción.

El modo más fácil para el neo-russelliano es acudir a los modos de presentación. Según Salmon (1986), por ejemplo, las creencias y los pensamientos son relaciones triádicas entre sujetos, proposiciones y modos de presentación[34]. Así, uno puede asentir a una proposición *p* cuando se le presenta bajo un modo de presentación, pero no asentir a la misma proposición cuando se le presenta bajo otro modo. Uno puede asentir a la proposición de que Venus es una estrella cuando Venus se le presenta bajo el modo "El primer cuerpo celeste visible por la tarde" pero no asentir al mismo tiempo a esa misma proposición cuando Venus se le presenta bajo el modo "El último cuerpo celeste visible por la mañana".

Ahora bien, el neo-russelliano niega que estos modos de presentación sean partes constituyentes de las proposiciones que expresamos al proferir enunciados que contienen nombres propios. Y en esto reside la mayor discrepancia con la corriente neo-fregeana; éstos sí afirman que ciertos modos de presentación o sentidos son partes

[34] No exactamente. Para Salmon, las creencias son predicados diádicos satisfechos por sujetos y proposiciones (y que atribuimos con enunciados como 'S cree que *p*'); la cuestión es que tras estos predicados diádicos se esconden unas relaciones triádicas. No hilaremos tan fino.

constituyentes de nuestros pensamientos singulares. Dentro de esta corriente, podemos identificar al menos dos tendencias. La primera admite que hay proposiciones singulares, proposiciones en parte constituidas por objetos particulares, aunque añade que sus modos de presentación también son partes constituyentes de estas proposiciones. La segunda desarrolla la idea de los sentidos "dependientes de objetos". Niega que los objetos particulares puedan ser partes constituyentes de las proposiciones, pero afirma que la identidad de un pensamiento singular depende de la identidad del objeto acerca del cual es. Evans (1982) es, seguramente, el principal valedor de esta hipótesis.

El grueso de *The Varieties of Reference* (1982) trata sobre cuáles son las condiciones necesarias y suficientes para que uno tenga un pensamiento singular. El texto es, además de extenso, arduo, y al igual que en las secciones anteriores con Putnam y Burge tampoco aquí tendremos ocasión de entrar a describir adecuadamente las ideas defendidas por Evans. Sólo esperamos que el siguiente bosquejo al menos sirva para caracterizar, aunque algo superficialmente, cierto posicionamiento externista de corte neo-fregeano.

Evans propone que las *Ideas* son aquellas herramientas conceptuales, conceptos de objetos, que posibilitan tener pensamientos singulares—un sujeto S tendrá un pensamiento singular *p* sobre un objeto *x* sólo si *p* está en parte constituido por una Idea de *x*. Es instanciando la misma Idea en diversos pensamientos que un sujeto puede pensar, en diferentes ocasiones, sobre un mismo objeto bajo el mismo modo.

> La Idea de un objeto es, pues, algo que posibilita al sujeto pensar acerca de un objeto en un número indefinido de pensamientos, tal que en cada uno de ellos estará pensando sobre el objeto del mismo modo.[35]

La cuestión es que, según el modelo que propone Evans (1982) sobre la naturaleza de las Ideas y sobre las condiciones necesarias y suficientes para su posesión, estas Ideas tienen cierta (digámoslo así) *naturaleza dual*; algunos elementos en su naturaleza hacen que sean herramientas indispensables para que a un sujeto le sea posible tener un

[35] An Idea of an object, then, is something which makes it possible for a subject to think of an object in a series of indefinitely many thoughts, in each of which he will be thinking of the object in the same way. (Evans (1982), p. 104)

pensamiento singular, otros que las propuestas de Evans se enmarquen dentro de un modelo más o menos fregeano.

Primero, para que S tenga una Idea I de *x*, es necesario que de hecho haya un *vínculo informativo*[36] entre *x* e I. La percepción de un objeto *x* posibilita que se forme un vínculo informativo entre *x* y el sujeto; mediante ese vínculo el sujeto recibe *desde x* información no-conceptual que va guardando en la Idea de ese objeto. Así, a la Idea le es esencial que haya cierta cadena causal-informativa entre el objeto del cual se origina la Idea, *x*, y la Idea misma—es imposible que una única Idea tenga más de un vínculo informativo, la identidad de la Idea viene en parte determinada por el objeto desde el cual recibe la información en cuestión. Por eso, es imposible individuar una Idea apelando a las propiedades internas del sujeto; qué objeto se encuentra en el origen del vínculo informativo en cuestión, qué objetos contiene el entorno del sujeto, es relevante para determinar qué Ideas y qué pensamientos singulares tiene.

Este vínculo informativo, esta cadena causal, determina cuál es el objeto al que *aspira* la Idea en cuestión, cuál es su *objetivo*. Por eso, el particular en el origen causal de una Idea I es el particular relevante para poder evaluar el valor de verdad de cualquier pensamiento singular en parte constituido por I—el que el pensamiento singular FA en parte constituido por la idea A sea verdadero o falso depende de que el objeto *x* que se encuentra en el origen causal de A tenga o no la propiedad predicada por el concepto F. Además, como hemos dicho, las identidades de la Idea y de los pensamientos singulares en parte constituidos por ella dependen también de ese particular en el origen del vínculo informativo. Ahora bien, ese objeto en el origen de la Idea ni constituye ni determina el contenido de la Idea. Expliquemos esto.

Según defiende Evans (1982), la existencia de uno de estos vínculos informativos no basta para que a un sujeto se le conceda que tiene una Idea de *x*; esta Idea ha de ser, en algún sentido, adecuada para *x*. Evans se adhiere a lo que él llama *Principio de Russell*:

> Russell sostuvo la opinión de que para poder estar pensando sobre un objeto o poder hacer un juicio sobre un objeto, uno debe *saber cuál* es el objeto en

[36] *Information-link*, en inglés.

> cuestión—uno debe *saber cuál* es el objeto sobre el cual está pensando uno. (Llamo a este principio el Principio de Russell).[37]

Para que alguien tenga un pensamiento singular sobre *x* no basta con que haya un vínculo informativo (o cualquier otro tipo de relación causal) entre ese sujeto y *x*, además uno tiene que saber que está pensando sobre *x*—uno no puede tener una Idea sobre *x* sin saber que esa Idea es sobre *x*. Pero para que el principio de Russell sea una tesis sustantiva, resulta necesario decir algo sobre qué quiere decir que uno sepa que su Idea es sobre *x*.

> Para hacer del Principio de Russell un principio sustancial, supondré que el conocimiento que requiere es lo que podría llamarse *conocimiento discriminatorio*: el sujeto tiene que tener la capacidad de distinguir el objeto de su juicio de todas las demás cosas. (...) Tenemos la idea de ciertas condiciones suficientes para ser capaz de discriminar un objeto de todas las demás cosas: por ejemplo, cuando uno puede percibirlo en el momento presente; cuando uno puede reconocerlo si se le presenta; y cuando uno conoce hechos distintivos sobre él.[38]

Uno puede tener pensamientos singulares sobre *x* sólo si puede discriminar entre *x* y cualquier otro objeto; hay al menos tres situaciones que permiten a S discriminar entre *x* y los demás objetos: cuando S está en una relación perceptiva (ostensiva) con *x*, cuando S tiene la capacidad de reconocer a *x* cuando se le presenta y cuando S sabe algún tipo de hecho distintivo sobre *x*. Al estar en alguna de estas tres situaciones, S adquiere un modo de identificación de *x*. Aquello que permite al sujeto discriminar entre *x* y los demás objetos, aquello que respondería a la pregunta de en virtud de qué sabe el sujeto que su pensamiento (o Idea) es sobre *x*, nos proporcionará el modo de identificación bajo el cual el sujeto piensa acerca de *x*. Y es ese modo el que constituye el contenido de la Idea, nos muestra bajo qué modo concibe S a *x* (cuando piensa sobre él empleando la Idea en cuestión), y explica en qué sentido puede discriminar S entre *x* y los demás objetos.

---

[37] Russell held the view that in order to be thinking about an object or to make a judgement about an object, one must *know which* object is in question—one must *know which* object it is that one is thinking about. (I call this principle Russell's Principle). (Evans (1982), p. 65)

[38] In order to make Russell's Principle a substantial principle, I shall suppose that the knowledge which it requires is what might be called *discriminating knowledge*: the subject must have a capacity to distinguish the object of his judgement from all other things. (...) We have the idea of certain sufficient conditions for being able to discriminate an object from all other things: for example, when one can perceive it at the present time; when one can recognize it if presented with it; and when one knows distinguishing facts about it. (Evans (1982), p. 89)

Así, pues, hay un modo de presentación que S relaciona con *x* y que constituye el contenido de una Idea que tiene S de *x*. Para determinar cuál es el contenido de una Idea en cuestión, no basta con identificar cuál es el objeto en el origen causal de esa Idea, es necesario presentar ese objeto de un modo que refleje el modo de presentación que constituye la Idea que toma parte en el pensamiento. Y, del mismo modo en que la relación causal que guardaba la Idea con cierto particular en su origen determinaba *el objetivo* de la Idea, el modo de presentación determina cuál es el *objeto* de la idea. Es debido (en parte) a estas cuestiones que la propuesta de Evans se enmarca dentro de un modelo neo-fregeano.

Hemos distinguido entre *el objetivo* de la Idea (el particular con el cual guarda un vínculo informativo) y *el objeto* de la Idea (el particular que satisface el modo que constituye el contenido de la Idea). Cuando el objetivo y el objeto de una Idea son el mismo particular *x*, entonces esa Idea permite al sujeto formar pensamientos singulares que refieren a *x*. Ahora, cabe la posibilidad de que el objetivo de la Idea no sea su objeto—que sea otro, que no haya ningún particular con el cual la Idea tenga un vínculo informativo, o que no haya ningún particular determinado por el modo de identificación que constituye el contenido de la Idea. Si esto fuera el caso, según Evans, el sujeto no tendría un pensamiento singular (a pesar de que, seguramente, creería tenerlo).

> No me parece que haya nada incoherente en la idea de que podría darse que, para un sujeto, fuera como si estuviera pensando sobre un objeto físico que (digamos) puede ver, y aún así que, precisamente porque no hay un objeto físico que está viendo, fallara en tener un pensamiento del tipo que supone que está teniendo. (...) La afirmación es simplemente que hay un tipo de pensamiento que tenemos algunas veces, típicamente expresado en la forma 'Este G es F', y podemos pretender tener un pensamiento de este tipo cuando, dada la ausencia de cualquier objeto apropiado, no hay ningún pensamiento de este tipo a tener.[39]

Por lo tanto, aunque los objetos en el origen causal de las Ideas no son partes constituyentes de esas Ideas, los modos de presentación (o sentidos) que conforman esas ideas sí dependen de esos objetos: si de hecho no hubiera objeto, no habría Idea (ni pensamiento singular), porque no habría modo de presentación *de nada*. Seguramente

[39] There does not seem to me to be anything incoherent in the idea that it may be, for a subject, exactly as though he were thinking about a physical object (say) which he can see, and yet that, precisely because there is no physical object he is seeing, he may fail to have a thought of the kind he supposes himself to have. (...) The claim is simply that there is a kind of thought we sometimes have, typically expressed in the form 'This G is F', and we may aim to have a thought of this kind when, in virtue of the absence of any appropriate object, there is no such thought to be had. (Evans (1982), pp. 45-46)

esta tesis de que hay sentidos dependientes de objetos sea el reflejo principal de cómo Evans conjuga un modelo externista singular con cierta tradición fregeana.

0.4. ESBOZO DE UN HIPOTÉTICO EXTERNISTA

Esperamos que todo lo dicho hasta ahora baste para una caracterización suficiente del modelo teórico contra el cual se dirigen los tres argumentos que estudiaremos. En dos palabras, la tesis esencial al externismo semántico sería la siguiente:

> **Externismo semántico:** Las propiedades internas de un sujeto no bastan para determinar en qué estado mental (intencional) se encuentra. Algunos factores externos son relevantes para determinar qué conceptos tiene ese sujeto, qué piensa, cree, desea o teme.

O, dicho de otro modo, el externismo semántico viene a ser la negación de la tesis internista que hemos presentado al comienzo de este capítulo. Entre los factores externos típicamente relevantes para determinar los conceptos y pensamientos de un sujeto están: los objetos particulares y las clases naturales presentes en el entorno del sujeto y la comunidad social (lingüística) en la que participa[40].

Por otro lado, las siguientes son afirmaciones que típicamente sostendrá el externista:

- Uno puede tener dos conceptos distintos C y C', saber que esos dos conceptos son distintos, pero relacionar el mismo *estereotipo* con ambos conceptos.
- Uno puede tener el concepto C que expresa mediante el término t sin saber cuál es la definición de t. Muchas veces, una comprensión *parcial* de un término t basta para adscribirle a uno competencia en cuanto a t y el concepto C.

[40] A pesar de que hemos diferenciado entre tres tipos de externismo dependiendo de qué factores externos se juzgue que son relevantes para la determinación de los estados mentales, lo más común es que un tipo de externismo esté "contaminado" por otro. Por ejemplo, Burge extiende el anti-individualismo que defiende a términos de clase natural (ahora bien, adecua la relevancia de las clases naturales al tipo de externismo que propone: es porque la comunidad lingüística en la que participa S usa un término de clase natural de cierto modo que la presencia de una clase u otra es relevante para determinar los estados mentales de S); y la tesis de la división de trabajo lingüístico de Putnam podría interpretarse como un préstamo tomado del externismo social, no difiere mucho (si algo) de la tesis de que podemos adquirir conceptos *deferencialmente*.

- Alguien puede tener un concepto C a pesar de que tiene creencias erróneas sobre qué criterios determinan qué cae (y qué no)dentro de la extensión de C. La descripción ϕ que S relaciona con C puede no bastar para determinar su referencia; lo que (al menos en parte) determina la referencia de un concepto es cierta cadena causal-comunicativa.
- Hay diferentes niveles de competencia lingüística. Uno puede ser un hablante competente del término t al adquirir ese término deferencialmente, porque su uso de t descansa en su participación en cierta comunidad lingüística.

Pero ésta es una caracterización de un externista *hipotético*; no pretendemos describir el posicionamiento de ningún autor en concreto. Uno tendrá que aceptar la tesis definitoria que hemos propuesto si es externista, pero no está en principio comprometido a las otras afirmaciones que hemos enumerado. Por ejemplo, uno puede decir que qué objetos particulares contiene el entorno de un sujeto es relevante para determinar sus estados mentales, pero negar que lo sea qué clases naturales contiene o la comunidad lingüística en la que participa; o uno podría negar alguna de las tesis que hemos expuesto arriba—por ejemplo, Evans (1982) niega que alguien pueda tener pensamientos singulares sobre un objeto *x* si la descripción que relaciona con el concepto que supuestamente refiere a *x* es satisfecha por un objeto *y* distinto a *x* (y defiende opiniones parecidas para los conceptos de clase natural).

Nuestra intención no era proponer una exposición exhaustiva de cuáles son las tesis y opiniones que adoptan distintos autores de corte externista, sino delimitar más o menos el posicionamiento teórico al que van dirigidos los argumentos que estudiamos en este trabajo. Hilaremos más fino cuando presentemos los argumentos.

Mencionemos algo antes de comenzar el grueso del trabajo. Cuando hemos descrito el modelo tradicional, hemos dicho que éste asume que el contenido es transparente—en un primer momento, parece que el externista está comprometido a negar que el contenido tenga esta propiedad. Hemos caracterizado al externista como proponiendo que uno puede tener el concepto AGUA sin saber qué es lo que de hecho determina que algo sea agua, y lo mismo se sigue por supuesto para el concepto BI-AGUA. En cuanto AGUA y BI-AGUA comparten estereotipo, parece en principio posible que alguien tenga estos dos conceptos y que falsamente crea que sólo tiene un único concepto (que

expresa al usar el término 'agua') donde realmente tiene dos; esto supondría un contraejemplo a la tesis de la transparencia de diferencia de contenido[41].

Es discutible que el externista esté comprometido a negar que el contenido es transparente[42]—nosotros creemos que lo está. Opinamos que, como hemos dicho, la posibilidad de que uno pueda tener un concepto C sin saber qué es lo que define ese concepto, y sin saber cuál es el criterio que determina qué cae (y qué no) dentro de la extensión de ese concepto, abre la puerta a escenarios donde uno tiene dos conceptos distintos pero cree que son el mismo—cree que sólo tiene un concepto donde de hecho tiene dos. A lo largo de este trabajo nos encontraremos con escenarios de este tipo, ya defenderemos entonces por qué opinamos que el externismo semántico es incompatible con la transparencia del contenido[43]. Aquí simplemente queríamos llamar la atención sobre el hecho de que al menos parece que el externista puede tener cierta dificultad en asumir que el contenido es transparente, ya que varios argumentos anti-externistas se basan (implícita o explícitamente) en este compromiso del externista. Concretamente, como veremos, así lo hace una de las versiones del primer argumento, y también el tercer argumento que estudiaremos.

Comencemos.

[41] La vertiente neo-russelliana del externismo semántico está también comprometida a rechazar la tesis de la trasnsparencia de mismidad de contenido. Según esta vertiente, los nombres propios *a* y *b* expresarán el mismo contenido si refieren al mismo objeto (ejemplos: 'Héspero' y 'Fósforo'; 'Tulio' y 'Cicerón'; 'Donostia' y 'San Sebastián'), pero uno puede no saber que de hecho tienen el mismo contenido. Así, se sigue que alguien puede tener, por ejemplo, los conceptos HÉSPERO y FÓSFORO sin saber que tienen el mismo contenido.

[42] Discutible y muy discutido; véanse, por ejemplo: Boghossian (1989, 1992a, 1994), Brown (2004), Burge (1988), Falvey y Owens (1994), Fodor (1998), Goldberg (1999, 2008), McLaughlin y Tye (1998), Owens (1989, 1990), Tye (1998).

[43] Adelantemos aquí nuestra posición: creemos que el único modo que tiene el externista de adherirse a la tesis de la transparencia de diferencia de contenido es apuntándose a la idea del *reemplazo conceptual*. Creemos que es preferible rechazar esta idea del reemplazo (sobre qué defiende el teórico del reemplazo: 2.3.3. de la primera parte y, sobre todo, 2.2. y 6.6. de la segunda parte; sobre cómo compatibilizar el externismo semántico con la transparencia del contenido acudiendo al reemplazo conceptual: 2.1. de la tercera parte; y sobre por qué creemos que el reemplazo no es una buena opción: 6.2.3. de la segunda parte).

(1).....

TRANSICIONES LENTAS, DISCRIMINACIÓN Y AUTO-CONOCIMIENTO

0. INTRODUCCIÓN

Supongamos que cierto individuo, Oscar, vive en la Tierra. Como sabemos, la Tierra está repleta de agua, cuya composición química es H_2O. Oscar tiene contacto a diario con el agua: bebe agua, se ducha con agua, de vez en cuando se baña en un lago lleno de agua, etcétera. Es más, tiene varias creencias acerca del agua, por ejemplo cree que el agua se hiela a cero grados, y además expresa varias de esas creencias profiriendo enunciados que contienen el término 'agua', como cuando profiere el enunciado 'el agua se hiela a cero grados'. Y supongamos también que hay algo que desconoce: supongamos que Oscar no sabe que el agua es H_2O, que desconoce cuál es la composición química del agua (quizás, lo más sencillo sea suponer que desconoce incluso los principios más rudimentarios de la química, y que nunca ha oído hablar de "composiciones químicas").

En el escenario que estamos describiendo existe cierto planeta más o menos distante a la Tierra, o cierto universo paralelo al que habita Oscar, al que llamaremos "Tierra Gemela". La Tierra Gemela es muy parecida a la Tierra: está habitada por humanos (los cuales llevan un estilo de vida muy parecido al que llevan los habitantes de la Tierra a principios del siglo XXI), contiene los mismos animales y las mismas plantas que la Tierra, tiene una atmósfera idéntica a la de la Tierra, etcétera. Pero también hay una

diferencia importante entre la Tierra y la Tierra Gemela: la Tierra Gemela carece de H_2O. En vez de H_2O, en la Tierra Gemela hay una sustancia cuya composición química transcribiremos como XYZ, y que denominaremos 'bi-agua'. La bi-agua es una sustancia muy parecida al agua: es transparente e insípida, se hiela a cero grados y hierve a cien. Los ríos y mares de la Tierra Gemela están llenos de bi-agua, cae bi-agua del cielo cuando llueve, y los habitantes de la Tierra Gemela beben bi-agua. Además, éstos hablan un idioma muy parecido al castellano (que denominaremos 'castellano gemelo'), y utilizan el término 'agua' para referirse a la bi-agua. Como es normal, los habitantes de la Tierra Gemela tienen creencias acerca de la bi-agua, las cuales expresan profiriendo enunciados en castellano gemelo que contienen el término 'agua'.

Resulta más bien plausible que, siguiendo la línea marcada por Putnam (1975) o Kripke (1980), asumamos que el agua y la bi-agua son dos sustancias necesariamente distintas. Por otro lado, el externista de clases naturales defiende que dos proferencias distintas del enunciado 'el agua es insípida', la primera hecha en la Tierra y la segunda en la Tierra Gemela, expresan proposiciones distintas, y que las creencias de los dos hablantes de la Tierra y la Tierra Gemela que motivaron las proferencias son de distinto tipo (porque al menos difieren en contenido)[44].

Sigamos con nuestra historia. Supongamos que cierto día, a pesar de que se acuesta en la Tierra, Oscar amanece en la Tierra Gemela, y que no se percata del cambio: si se le preguntara, respondería que siempre ha habitado el mismo planeta. Continúa con su vida en la Tierra Gemela, y entra en contacto con la bi-agua: bebe bi-agua, se baña en lagos repletos de bi-agua, se ducha con bi-agua, y utiliza bi-agua para regar las plantas de su balcón. Oscar nunca descubre que ha cambiado de planeta, y nunca llega a conocer que la bi-agua, el líquido que bebe a diario desde que se trasladó a la Tierra Gemela, tiene la composición química XYZ. Pasa el tiempo en la Tierra Gemela, el suficiente como para que Oscar adquiera el concepto BI-AGUA y aprenda a hablar castellano gemelo; comienza a tener creencias en parte constituidas por el concepto BI-AGUA, y empieza a proferir enunciados de castellano gemelo que son sobre la bi-agua. Pasan los años y Oscar cambia varias veces de tierra, sin percatarse en ningún momento

[44] Mencionemos aquí, aunque sólo sea para evitar equívocos, que, a pesar de que las aportaciones de Kripke (1980) y Putnam (1975) resultan esenciales para el desarrollo de las teorías externistas de clases naturales, ninguno de los dos defiende explícitamente una teoría externista para los contenidos mentales (como vimos en la introducción, Putnam (1975) la rechaza explícitamente).

de que tales transiciones están teniendo lugar, y pasando cada vez el tiempo suficiente en el nuevo escenario para adquirir o comenzar a usar el concepto apropiado para ese escenario. Cierto día, en la Tierra, frente a un lago de agua helada, Oscar piensa un pensamiento, el cual expresaría profiriendo el enunciado 'el agua se hiela a cero grados'. Fin de la historia.

El argumento incompatibilista que presentaremos en esta primera parte propone escenarios como el descrito arriba, comúnmente denominados 'escenarios de *transición lenta*', donde un sujeto viaja de un entorno a otro varias veces (sin darse cuenta de ello en ningún momento), pasa el tiempo suficiente en la nueva comunidad lingüística como para aprender el nuevo idioma y adquirir un concepto nuevo, y tiene un pensamiento en parte constituido por el concepto que acaba de adquirir. Brevemente, el argumento defiende que, por ejemplo, Oscar no puede conocer que está pensando que el agua se hiela a cero grados, ya que podría estar pensando que la bi-agua se hiela a cero grados y, al ser las dos situaciones "internamente indistinguibles", carece de evidencia introspectiva suficiente para excluir ese escenario. Fue Paul Boghossian (1989) quien por primera vez presentó el argumento en un modo elaborado, y recientemente Jessica Brown (2004) ha propuesto una nueva versión del mismo argumento.

El objetivo de esta parte es presentar y discutir el argumento que acabamos de mencionar. Empezaremos por exponer las dos versiones del argumento que estudiaremos (la de Boghossian (1989) y la de Brown (2004)), y presentaremos algunas de las respuestas ofrecidas al argumento (Burge (1988), Falvey y Owens (1994), McLaughlin y Tye (1998) y Brown (2004)). Evaluaremos las distintas respuestas presentadas (así como algunas críticas hechas a las respuestas), y defenderemos, primero, que hay al menos cierta tensión entre dos de las premisas y presuposiciones en las que se basa el argumento y, segundo, que Falvey y Owens al menos marcan la dirección adecuada para una respuesta compatibilista acertada.

1. EL ARGUMENTO SEGÚN BOGHOSSIAN

El objetivo de Boghossian en "Content and Self-Knowledge" (1989) es, según él mismo, demostrar que explicar la naturaleza del auto-conocimiento resulta problemático (al menos si aceptamos ciertas tesis acerca de la naturaleza de nuestros estados mentales intencionales). Presenta tres modelos alternativos de auto-conocimiento, y mantiene que el segundo resulta incompatible con cualquier teoría *relacionista* del contenido mental, entre las cuales se encuentra la teoría externista (es éste el argumento incompatibilista que estudiaremos). Como, de acuerdo con él, los otros dos modelos no son aceptables, concluye que hay un problema para explicar la naturaleza del auto-conocimiento. En el artículo también propone un segundo argumento incompatibilista, *el argumento de la memoria*, que estudiaremos en la segunda parte; por ahora nos centraremos sólo en el primero. Empezaremos caracterizando brevemente los tres modelos de auto-conocimiento que menciona: auto-conocimiento inferido de otras creencias, auto-conocimiento basado en observaciones internas y auto-conocimiento basado en nada; luego haremos explícito cada uno de los pasos que da el argumento incompatibilista que propone.

1.1. TRES MODELOS DE AUTO-CONOCIMIENTO

1.1.1. Auto-conocimiento inferido de otras creencias.

Este primer modelo propone que adquirimos conocimiento de nuestros propios estados mentales infiriéndolo de otras creencias acerca de nuestra conducta. Esto es, cómo conocemos en qué estado mental nos encontramos no difiere mucho de cómo conocemos en qué estado mental se encuentran los demás. Parece obvio que el único modo que tenemos de saber en qué estado mental se encuentran terceras personas es observando su conducta; por ejemplo, si el padre de Elizabeth observa que ésta se comporta de forma nerviosa cada vez que Darcey se encuentra cerca, podrá inferir (y quizás saber) que Elizabeth está enamorada de Darcey; o, por otro lado, si Elizabeth observa que Darcey profiere el enunciado 'Los habitantes de la campiña son gente poco sutil', podrá inferir que Darcey cree que los habitantes de la campiña son gente poco sutil.

Según esta primera alternativa, el auto-conocimiento sigue exactamente el mismo procedimiento: conocemos en qué estado mental nos encontramos observando nuestra propia conducta. Si Coraline observa que se agacha y pone la mano en su barriga, podrá inferir de tal observación (y así saber) que siente dolor de barriga; si observa que agarra una bufanda y sus katiuskas amarillas antes de salir de casa, Coraline podrá inferir de tal observación (y saber) que cree que hace frío y llueve.

Brevemente, la tesis principal de este primer modelo es que el auto-conocimiento se basa en observaciones sobre nuestra conducta: evidencia pública, observable, a la cual nadie tiene ningún tipo de acceso *privilegiado*. En una discusión entre Coraline y sus padres acerca de qué es lo que cree o piensa Coraline, o acerca de si le duele la barriga o no, ésta no goza de ningún tipo de autoridad en comparación con sus padres, ya que tanto sus opiniones sobre sus creencias así como las opiniones de sus padres se basan en una evidencia a la cual todos tienen el mismo tipo de acceso. Esto es, este primer modelo no es un modelo de auto-conocimiento *autoritativo*: niega que tengamos ningún tipo de acceso privilegiado a nuestros estados mentales.

1.1.2. Auto-conocimiento basado en "observaciones internas".

Lo que caracteriza principalmente el segundo modelo es cierta analogía o metáfora; la metáfora viene a decir que nuestro auto-conocimiento se basa en "percepciones internas". Del mismo modo en que nuestros sentidos nos proporcionan información acerca de sucesos y objetos externos a nosotros, el "modelo observacional del auto-conocimiento" defiende que hay cierto tipo de "sentido interno" que nos proporciona información acerca de los sucesos y estados en nuestras propias mentes; *inferimos* nuestras creencias empíricas de las percepciones que tenemos, y nuestras creencias sobre nuestros estados mentales de algún tipo de "percepción interna". Es más o menos común utilizar la metáfora del "ojo interno" que "percibe" nuestros estados mentales para explicar este modelo; así lo hace, por ejemplo, Tyler Burge (1979):

> Un modelo que asemeja la relación entre una persona y los contenidos de su pensamiento a ver, donde se entiende que ver es un tipo de experiencia inmediata, directa.[45]

Acudiendo a otra metáfora recurrente, Davidson (1987) equipara al auto-conocimiento según este modelo a un espectador que observa una obra que se representa en un teatro:

> En una versión tosca, pero familiar, [esta caracterización de la mente] dice así: la mente es un teatro en el cual el yo consciente observa un show que se está ofreciendo (las sombras en la pared). El show consiste en "apariencias", datos de los sentidos, qualia, lo que es dado en la experiencia. Lo que aparece en el escenario no son los objetos ordinarios del mundo que el ojo externo registra y el corazón quiere, sino los que se suponen sus representantes.[46]

Bien, está más o menos claro cuál es la metáfora pero, protestan algunos filósofos, no está para nada claro cómo debemos entenderla. Si alguien propone una analogía entre el auto-conocimiento y la percepción, debería al menos explicar en qué consiste esta analogía. Shoemaker (1994) intenta caracterizar lo que sería este modelo de auto-conocimiento, para luego concluir que no es aceptable. Enumera las características más significativas de la percepción de objetos externos, y señala que, seguramente,

---

[45] ...a model that likens the relation between a person and the contents of his thought to seeing, where seeing is taken to be a kind of direct, immediate experience. (Burge (1979), p. 133)

[46] In one crude, but familiar, version, [this picture of the mind] goes like this: the mind is a theater in which the conscious self watches a passing show (the shadows on the wall). The show consists of 'appearances', sense data, qualia, what is given in experience. What appear on the stage are not the ordinary objects of the world that the outer eye registers and the heart loves, but their purported representatives. (Davidson (1987), p. 105)

podríamos caracterizar el modelo observacional como manteniendo que el auto-conocimiento de algún modo tiene cuatro de esas características:

> Aunque la percepción sensorial provee a uno con percatación de hechos, esto es, percatación *de que* tal-y-cual es el caso, lo hace mediante la percatación de objetos. La percatación de hechos se explica mediante la percatación de uno de los objetos envueltos en esos hechos. (...)
> La percepción sensorial aporta "información de identificación" sobre el objeto de la percepción. Cuando percibe uno tiene la capacidad de captar un objeto de entre otros, distinguiéndolo de los otros mediante información, dada por la percepción, sobre sus propiedades tanto relacionales como no-relacionales. (...)
> La percepción de objetos conlleva la percepción de sus propiedades intrínsecas, no-relacionales. Podemos percibir relaciones entre objetos que percibimos; pero en ningún modo percibiríamos estos objetos, y así no podríamos percibir las relaciones entre ellos, si no se presentaran como teniendo propiedades intrínsecas, no-relacionales. (...)
> Los objetos de la percepción son objetos potenciales de atención. Sin cambiar lo que percibe uno, uno puede cambiar la atención de un objeto percibido a otro, aumentando así la capacidad de uno de adquirir información sobre él.[47]

Según la caracterización de Shoemaker, pues, el modelo observacional diría que guardamos cierto tipo de relación con nuestros estados mentales, en cierto modo análoga a la percepción de objetos externos, que hace que nos percatemos de estos estados mentales (y descubrimos que estamos en el estado M porque descubrimos, nos percatamos de, el estado M), que nos aporta "información de identificación" de esos estados, distinguiéndolos de otros, que sólo nos aporta conocimiento de las propiedades intrínsecas de esos estados (conoceríamos sus propiedades relacionales sólo infiriéndolos de sus propiedades intrínsecas), y esta "percepción" sería distinta a la atención que podríamos poner en los estados mentales que "percibimos". Shoemaker niega que un modelo tal sea sostenible (no mencionaremos aquí los motivos).

Hay estados característicamente fenoménicos: padecer dolor, percibir rojo (sólido o salado), sentir tristeza. En principio, y sin entrar en mayores caracterizaciones, parece que es relativamente sencillo proponer un modelo observacional de auto-conocimiento

[47] While sense perception provides one with awareness of facts, i.e., awareness *that* so and so is the case, it does this by means of awareness of objects. One's awareness of the facts is explained by one's awareness of the objects involved in these facts. (...) Sense perception affords "identification information" about the object of perception. When one perceives one is able to pick out one object from others, distinguishing it from the others by information, provided by the perception, about both its relational and its nonrelational properties. (...)The perception of objects standardly involves perception of their intrinsic, nonrelational properties. We can perceive relations between things we perceive; but we wouldn't perceive these things at all, and so couldn't perceive relations between them, if they didn't present themselves as having intrinsic, nonrelational properties. (...) Objects of perception are potential objects of attention. Without changing what one perceives, one can shift one's attention from one perceived object to another, thereby enhancing one's ability to gain information about it. (Shoemaker (1994), pp. 252-253)

de tales estados[48]. Estos estados tienen propiedades fenoménicas, padecimiento de dolor, percepción de rojo (sólido o salado), sentimiento de tristeza, a las que podemos llamar '*qualia*'. Uno "percibe" estos *qualia* mediante introspección y, sobre la base de la evidencia que obtiene, puede llegar a conocer que está en un estado tal. Ahora, este modelo parece más difícil de aplicar a estados característicamente intencionales, como juzgar que *p*, pensar que *p*, creer que *p* o temer que *p*—no está claro que estos estados tengan propiedades fenoménicas, que tengan *qualia*. Burge (1996) acude a nociones de carácter epistemológico para intentar caracterizar lo que sería el auto-conocimiento de estados intencionales dentro de un modelo observacional:

> El modelo no necesita proponer ninguna presentación fenoménica en el auto-conocimiento, aunque dejar de lado tal afirmación debilita la analogía con la observación. La tesis fundamental es que la justificación epistémica que tiene uno para el auto-conocimiento siempre descansa en parte en la existencia de un patrón de relaciones verídicas, aunque brutas, contingentes y no-racionales – los cuales plausiblemente son siempre relaciones causales – entre el objeto del auto-conocimiento (las actitudes que se revisan) y los juicios sobre las actitudes.[49]

Tal y como sugiere esta cita, una característica importante del modelo observacional (característica que, por ejemplo, la distingue del tercer modelo que presentaremos ahora) es que abre la puerta a errores *brutos*. Cuando uno descubre mediante observación empírica un hecho externo, aquello que descubre podría ser falso—los escenarios escépticos son siempre en principio posibles en estos casos (porque uno podría estar alucinando, o porque los órganos perceptivos de uno podrían funcionar incorrectamente). Parece razonable pensar que la posibilidad del error bruto le es esencial al conocimiento obtenido mediante observación. Por ello, si el auto-conocimiento se basa en algún tipo de observación, parecería que el error bruto es siempre en principio posible[50].

Resumiendo, este modelo defiende que nuestro conocimiento de nuestros propios estados mentales se basa en evidencia, pero que esta evidencia tiene cierto carácter privado, subjetivo, y sobre la base de esa evidencia el sujeto *infiere* que se encuentra en

---

[48] Por supuesto, no estamos sugiriendo que la caracterización que mencionamos sea aceptable.

[49] The model need not claim any phenomenological presentation in self-knowledge, though waiving such a claim weakens the analogy to observation. The fundamental claim is that one's epistemic warrant for self-knowledge always rests partly on the existence of a pattern of veridical, but brute, contingent, non-rational relations – which are plausibly always causal relations – between the subject matter (the attitudes under review) and the judgments about the attitudes. (Burge (1996), p. 253)

[50] Más acerca de las opiniones de Burge (1996) sobre la posibilidad del error bruto en el auto-conocimiento en los últimos párrafos de la sección 3.3. de esta primera parte.

un estado u otro. El sujeto goza de cierta capacidad que le proporciona de modo *inmediato* información acerca de sus estados mentales, capacidad de la que carecen terceras personas (acerca de los estados mentales del primero, se entiende). Es éste, pues, un modelo de auto-conocimiento autoritativo, defiende que el sujeto tiene cierto acceso privilegiado a sus propios estados mentales.

1.1.3. Auto-conocimiento basado en nada.

La tercera opción presentada por Boghossian mantiene que el auto-conocimiento no se basa en ningún tipo de evidencia; el sujeto simplemente conoce en qué estado mental se encuentra, sin tener que inferirlo de evidencia alguna, sea ésta interna o externa; se asemejaría por eso al conocimiento de otras proposiciones (algunas de ellas contingentes) que no se basa en ningún tipo de observación o evidencia. En palabras de Boghossian, el conocimiento empírico ordinario supone un *logro cognitivo*, y se basa en una *epistemología sustancial*, al contrario del conocimiento de proposiciones contingentes que no se basa en evidencia alguna. Conocer que yo estoy aquí ahora, por ejemplo, no me supone ningún logro cognitivo, y no me es preciso ningún tipo de evidencia para obtener tal conocimiento. Todo pensamiento del tipo YO ESTOY AQUÍ AHORA es veraz y justificado, y basta con hacer un juicio de ese tipo para poder obtener tal conocimiento inmediatamente. Lo mismo sucedería, según este tercer modelo, con el auto-conocimiento—basta con que un sujeto haga un juicio del tipo YO ESTOY EN EL ESTADO MENTAL M para que inmediatamente conozca que se encuentra en el estado mental M. De algún modo, estos juicios son siempre verdaderos, simplemente no se puede dar el caso de que uno juzgue que está en el estado mental M y que no se encuentre en el estado mental M, no hay espacio para errores *brutos*.

> De acuerdo con algunos filósofos, algunos juicios acerca de uno mismo son esencialmente auto-verificantes. Previo al juicio de que estoy celoso, por ejemplo, puede que no haya nada que pueda verificar si lo estoy; pero el pensarlo hace que lo haya. El juicio de que estoy celoso, una vez hecho, es, por ello, tanto verdadero como justificado. Pero, de nuevo, no se requiere de ninguna evidencia para el juicio.[51]

[51] According to some philosophers, certain self-regarding judgments are essentially self-verifying. Antecedent to the judgment that I am jealous, for example, there may be no fact of the matter about whether I am; but thinking it makes it so. The judgment that I am jealous, when made, is, therefore, both true and justified. But, again, no evidence is required for the judgment (Boghossian (1989), pp. 165-166)

Siguiendo con Boghossian, según este modelo algunos estados mentales (si no todos) son tal que dependen de los juicios del sujeto para su existencia: si el sujeto no juzga que está celoso, entonces el sujeto no está celoso.

Resumiendo, el tercer modelo de auto-conocimiento defiende que el auto-conocimiento no se basa en ningún tipo de evidencia, que no conocemos en qué estado mental nos encontramos infiriéndolo de ningún tipo de observación (ni siquiera de "observaciones internas"). Y como a un sujeto le es suficiente juzgar que está en el estado mental M para conocer que está en el estado mental M (por contra, juzgar que una tercera persona se encuentra en el estado mental M no basta para que esa persona se encuentre en el estado mental M), también este tercer modelo es un modelo de auto-conocimiento autoritativo.

1.1.4. Breves notas sobre los tres modelos.

Ésos, pues, los tres modelos de auto-conocimiento que presenta Boghossian. Éste protesta que el primer y el tercer modelo son problemáticos de por sí, y propone un argumento incompatibilista para el segundo modelo, que presentaremos en la siguiente sección y sobre el cual versa toda esta primera parte. Tal y como dice Boghossian, el modelo basado en inferencia resulta completamente anti-intuitivo, parece un hecho obvio que sí tenemos acceso privilegiado a muchos de nuestros estados mentales y, además, es el auto-conocimiento *autoritativo* el que supuestamente amenaza el externismo semántico. No entraremos a discutir este modelo por lo tanto.

Más (mucho más) se debería decir sobre el modelo observacional, pero no es éste el lugar para ello. El argumento de Boghossian viene a decir que es éste el modelo que resulta incompatible con el externismo semántico, así que intentaremos quedarnos con lo esencial para poder presentar y discutir el argumento. Como hemos dicho, más allá de la metáfora, no está claro en qué consiste este modelo—ahora, las siguientes ideas, si no llegan a caracterizarlo completamente, sí le son esenciales. Primero, el modelo observacional defiende que el auto-conocimientos se basa en evidencia que tiene cierta naturaleza interna (sólo accesible por lo tanto al sujeto en cuestión), de la cual el sujeto *infiere* en qué estado mental se encuentra. Además, también dice que los mismos principios de naturaleza epistémica que rigen el conocimiento de proposiciones

empíricas rigen el auto-conocimiento de proposiciones sobre nuestros estados mentales. Si hubiera alguna norma que impusiera condiciones al conocimiento que tiene S de que *p* (siendo *p* una proposición empírica), entonces la misma norma impondría condiciones análogas al conocimiento que tiene S de que *q* (donde *q* es una proposición sobre en qué estado mental se encuentra S).

Por último, no creemos que la caracterización que hace Boghossian del tercer modelo sea suficiente (tampoco diremos nosotros mucho); más adelante, cuando discutamos las críticas que hace Boghossian al modelo de auto-conocimiento basado en los pensamientos *cogito* de Burge (sección 3.3.), diremos alguna cosa sobre este modelo de "auto-conocimiento basado en nada".

Sólo mencionemos aquí que este modelo, *tal y como lo ha descrito Boghossian*, no es aceptable, ya que tiene la consecuencia de que el auto-conocimiento no es falible (de acuerdo con la caracterización del modelo que ha hecho Boghossian, a uno le basta con pensar un juicio del tipo ESTOY EN EL ESTADO MENTAL M para saber que está en el estado mental M). Y, de hecho, el auto-conocimiento sí es falible, uno puede tener creencias falsas sobre en qué estado mental se encuentra. Adelantemos también aquí que, no obstante, sí creemos que hay algunos espacios donde el auto-conocimiento no es falible, y donde hacer un juicio ESTOY EN EL ESTADO MENTAL M basta para que uno esté en el estado mental M (pero trataremos estas cuestiones con más detenimiento en el capítulo 3 concerniente a Burge, especialmente en la sección 3.3.).

Terminemos esta sección sobre los tres modelos diciendo que no es el objetivo de este trabajo concluir proponiendo un modelo adecuado y plausible de auto-conocimiento autoritativo que sea compatible con la semántica externista. Lo que nos proponemos en los capítulos siguientes es solamente, basándonos en las opiniones de distintos autores, defender que el externismo es compatible con el auto-conocimiento autoritativo; en ningún momento intentaremos ofrecer una caracterización completa y satisfactoria de este fenómeno. Digamos aquí sólo que, aceptando algunas propuestas de Burge (1988), juzgamos que el externismo semántico ni siquiera puede suponer una amenaza para el auto-conocimiento autoritativo; si lo hiciera, si planteara algún tipo de amenaza a nuestro auto-conocimiento, lo haría al amenazar el conocimiento autoritativo que tenemos de los *contenidos* de nuestros estados mentales (ya que el externismo

semántico es una teoría sobre qué contenidos tienen nuestros estados). La cosa es que, con Burge (1988), opinamos que podemos *blindar* el conocimiento que tenemos de esos contenidos. Por lo tanto se sigue, por un lado, que ninguna teoría semántica (tampoco la externista) puede amenazar *ese* conocimiento que tenemos (el conocimiento de los contenidos de nuestros estados) y, por el otro, que cualquiera que sea el modelo adecuado de auto-conocimiento autoritativo, será compatible con la semántica externista.

Ya trataremos estos temas con más detenimiento en el capítulo 3; comencemos presentando el argumento.

1.2. EL ARGUMENTO

Boghossian comienza esbozando una primera versión del argumento, la cual sólo iría en contra del externismo semántico (no contra toda teoría semántica *relacionista*) y que descarta más adelante. Según esta primera versión, para poder saber que estoy pensando sobre el agua, debo saber primero que mi pensamiento contiene el concepto AGUA y no el concepto BI-AGUA. Pero para poder saber tal cosa tendría que saber primero que mis pensamientos son típicamente sobre H_2O, y no sobre XYZ, y es evidente que sólo puedo saber esto mediante investigación empírica. Brevemente, esta primera versión del argumento viene a decir que, si el externismo es verdadero, no podemos conocer nuestros pensamientos mediante introspección, porque sus contenidos dependen de factores externos que es imposible que conozcamos mediante introspección. Pero esta primera versión no es sostenible. El mismo Boghossian[52] señala que el argumento se basa en un principio inaceptable:

> (Nec) Si el que se dé *q* es una condición necesaria para que S conozca que *p*, entonces S ha de conocer que se da *q* para poder conocer que *p*.

[52] "Esta línea de argumentación es sin ninguna duda demasiado precipitada. (...) resulta controvertido, por ponerlo de forma suave, si [para conocer que *p* uno] necesita conocer todas las condiciones que hacen tal conocimiento posible" ("This line of reasoning is no doubt too swift. (...) it is controversial, to put it mildly, whether [in order to know that *p* one] needs to know all the conditions that make such knowledge possible." (Boghossian (1989), pp. 165-166))

(Nec) es falso, propone unas condiciones demasiado exigentes para el conocimiento. Supongamos, por ejemplo, que Maider cree (verazmente) que Isabella Rossellini es una de las protagonistas de *Blue Velvet.* Supongamos que lo cree sobre una evidencia que a todas luces parece suficiente (ha visto la película varias veces leyendo en los títulos de crédito que Isabella Rossellini es una de las actrices protagonistas y, ya que ha visto varias películas de la actriz así como muchas fotografías suyas, es capaz de reconocerla cuando la ve en una imagen); querríamos decir que Maider *sabe* que Isabella Rossellini es una de las protagonistas de *Blue Velvet*. Pero supongamos también que Maider nunca ha oído hablar de Ingrid Bergman (madre de Isabella Rossellini), que no sabe que Ingrid Bergman existió. Aceptando cierta tesis sobre la necesidad del origen, el que Ingrid Bergman haya existido resulta una condición necesaria para que Maider sepa que Isabella Rossellini es una de las protagonistas de *Blue Velvet*. Dado que Maider no sabe que Ingrid Bergman existió, se sigue sobre la base de (Nec) que Maider no sabe que Isabella Rossellini es la protagonista de *Blue Velvet*. Es más, Maider no podrá conocer nada acerca de Isabella Rossellini hasta que descubra que Ingrid Bergman existió. La conclusión resulta demasiado escéptica, (Nec) nos lleva a un escenario que resulta demasiado escéptico para ser deseable.

Por ello, Boghossian propone un principio más plausible sobre el cual erigir su argumento: "Parece que el concepto ordinario de conocimiento no exige más que la exclusión de las hipótesis alternativas "relevantes" (como sea que haya que entender esto exactamente); y la mera posibilidad lógica no confiere tal relevancia."[53] Esto es, parece que para Boghossian el siguiente principio sí es verdadero:

(RA) Si
(i) *q* es una alternativa relevante a *p*, y
(ii) la creencia de S de que *p* está basada en evidencia que es compatible con que sea el caso que *q*, entonces

S no sabe que *p*.[54, 55]

[53] The ordinary concept of knowledge appears to call for no more than the exclusion of "relevant" alternative hypotheses (however exactly that is to be understood); and mere logical possibility does not confer such relevance. (Boghossian (1989), pp. 165-166)
[54] (RA) If (i) *q* is a relevant alternative to *p*, and (ii) S's belief that *p* is based on evidence that is compatible with its being the case that *q*, then S does not know that *p*. (Falvey y Owens (1994), p. 116)
[55] Esta formulación se debe a Falvey y Owens (1994), más adelante en el cuarto capítulo presentaremos y discutiremos la respuesta compatibilista que proponen.

Pero (RA), por sí solo, no demuestra que el externismo y el auto-conocimiento autoritativo resultan incompatibles; incluso si la evidencia introspectiva que obtengo cuando pienso que el agua es insípida es compatible con un escenario en el que estoy pensando que la bi-agua es insípida, no se sigue sobre (RA) que no puedo conocer que estoy pensando que el agua es insípida, ya que el escenario alternativo (en el que estoy pensando que la bi-agua es insípida) no es relevante (y la primera condición impuesta por el antecedente de (RA) es que el escenario alternativo ha de ser *relevante*). Es con el fin de hacer tales escenarios relevantes que Boghossian introdujo un escenario de transición lenta.

Recordemos el caso de Oscar. Éste sufre varias transiciones lentas entre la Tierra y la Tierra Gemela, cada vez pasa el tiempo suficiente para adquirir el concepto en cuestión, Oscar nunca descubre que alguna vez ha cambiado de entorno, y desconoce tanto la composición química del agua como la de la bi-agua. Cierto día en la Tierra, frente a un lago de agua, Oscar tiene un pensamiento, que expresaría profiriendo el enunciado 'el agua se hiela a cero grados'. Dado que se encuentra en la Tierra, concluimos, sobre la base de la teoría externista, que Oscar está pensando que el agua se hiela a cero grados. ¿Puede conocer Oscar sin ayuda de investigación empírica, mediante introspección, que está pensando que el agua se hiela a cero grados? Según Boghossian, no; he aquí el argumento:

Debido a las transiciones que ha sufrido a lo largo de su vida, el que se encuentre en la Tierra Gemela pensando que la bi-agua se hiela a cero grados es un escenario relevante para Oscar (un escenario relevante en el que es falso que está pensando que el agua se hiela a cero grados). Así, y sobre la base de (RA), tendrá que excluir que se encuentra en el escenario alternativo para poder conocer que está pensando que el agua se hiela a cero grados. Recordemos que estamos suponiendo que conocemos en qué estado mental nos encontramos sobre la base de evidencia que adquirimos gracias a ciertas "observaciones internas". El estado interno actual de Oscar es indistinguible del estado interno en el que se encontraría si estuviera en la Tierra Gemela pensando que la bi-agua se hiela a cero grados: la evidencia que obtiene mediante introspección no es suficiente como para excluir el escenario relevante, es compatible con ese escenario.

Por lo tanto, concluye el argumento de Boghossian, Oscar no puede conocer sólo mediante introspección que está pensando que el agua se hiela a cero grados.[56]

He aquí, esquematizado brevemente, la estructura del argumento:

1. (RA) Si (i) *q* es una alternativa relevante a *p*, y (ii) la creencia de S de que *p* está basada en evidencia que es compatible con que sea el caso que *q*, entonces S no sabe que *p*.
2. S puede conocer en qué estado mental se encuentra mediante inferencias desde evidencia que obtiene "observando" sus estados internos.
3. Para Oscar, el que se encuentre en la Tierra Gemela pensando que la bi-agua se hiela a cero grados es una alternativa relevante en la que es falso que está pensando que el agua se hiela a cero grados.
4. La evidencia que obtiene Oscar mediante introspección es la misma en el escenario actual y el escenario relevante, ya que si estuviera en la Tierra Gemela estaría en el mismo estado interno que en el escenario actual.
5. Por lo tanto, la evidencia que obtiene Oscar mediante introspección es compatible tanto con el escenario actual en el que está pensando que el agua se hiela a cero grados como con el escenario alternativo relevante en el que piensa que la bi-agua se hiela a cero grados.
6. Por lo tanto, sobre la base de (RA), Oscar no puede conocer, sólo por introspección, que está pensando que el agua se hiela a cero grados.

Nos gustaría mencionar un par de características del argumento. Primero, creemos necesario resaltar cuáles son el objetivo y el alcance real del argumento (tenemos la sensación de que algunos autores no tienen en cuenta estos puntos cuando proponen sus respuestas). Por de pronto, no podemos olvidar que el objetivo de Boghossian no es demostrar que el externismo semántico resulta incompatible con cualquier modelo de auto-conocimiento autoritativo; al contrario, el argumento pretende demostrar que el externismo semántico es incompatible con cierto modelo concreto de auto-conocimiento

[56] El argumento tal y como lo hemos presentado tiene como objetivo sólo a las teorías externistas, pero ya hemos mencionado que Boghossian pretende extenderlo a toda teoría *relacionista* del contenido mental. La importancia del argumento se hace más notoria si, con Boghossian, aceptamos que toda teoría semántica aceptable habrá de ser relacionista, independientemente de que sea internista o externista. En el artículo, defiende que una teoría funcionalista del contenido caería ante el mismo argumento, porque dentro del modelo observacional del auto-conocimiento la introspección sólo nos da información sobre las propiedades intrínsecas de nuestros estados mentales, y las propiedades causales no sobrevienen en las propiedades intrínsecas. Pero dejaremos de lado la efectividad del argumento contra estas teorías; al fin y al cabo, lo que nos interesa es si el externismo semántico supone algún tipo de amenaza para el auto-conocimiento autoritativo.

autoritativo. En principio, una posible estrategia para defender que el externismo es compatible con el auto-conocimiento autoritativo sería rechazar ese modelo concreto.

Por otro lado, al menos no queda claro que la conclusión del argumento sea que si el modelo observacional del auto-conocimiento es verdadero entonces nosotros no tenemos acceso privilegiado a nuestros estados mentales. La conclusión del argumento es que Oscar, víctima de una transición lenta, no tiene acceso privilegiado a sus propios estados mentales; en principio, sólo la autoridad de aquéllos que son víctimas de transiciones lentas resulta amenazada según el argumento (más adelante, cuando en el capítulo 6 discutamos la propuesta compatibilista de Brown, diremos más sobre esto).

Hasta aquí la presentación del argumento de Boghossian. En el siguiente capítulo presentaremos la versión del argumento defendida por Brown (2004), y en los siguientes, las que creemos son las respuestas más relevantes al argumento.

2. EL ARGUMENTO SEGÚN BROWN

En *Anti-individualism and Knowledge* (2004), Jessica Brown trata varios temas de carácter epistemológico relacionados con el externismo semántico; entre otras cuestiones, presenta y discute algunos argumentos incompatibilistas. Presenta primero un argumento que llama *argumento de la discriminación*; el argumento tal y como lo formula Brown no es exactamente el mismo que hemos presentado en el capítulo anterior pero, dados sus rasgos comunes, creemos que podemos entender que Boghossian (1989) y Brown (2004) presentan dos versiones distintas del mismo argumento. Opinamos que presentarla aquí merece la pena por al menos dos motivos: primero, el argumento se basa explícitamente en un principio epistémico que relaciona capacidades discriminatorias y conocimiento, así como en el supuesto compromiso del externista a negar que el contenido es transparente; segundo, Brown defiende que su versión del argumento es inmune a las respuestas de Burge (1988), Falvey y Owens (1994) y McLaughlin y Tye (1998) al argumento de Boghossian.

2.1. EL ARGUMENTO

El argumento de la discriminación se basa en casos de transición lenta como el presentado al principio de esta parte. Muy brevemente, el argumento viene a decir que, si el externismo semántico fuera verdadero, un sujeto que fuera víctima de una transición así no podría saber sólo mediante introspección qué está pensando, ya que no podría distinguir entre la situación actual y una situación contrafáctica en la que estaría pensando otra cosa, y tales capacidades discriminatorias son necesarias para el conocimiento:

> Según el argumento de la discriminación, el anti-individualismo amenaza el acceso privilegiado minando la habilidad de un sujeto de distinguir a priori entre los contenidos de pensamiento que tiene en la actualidad y los contenidos de pensamiento que hubiera tenido en varias situaciones contrafácticas.[57]

En dos palabras: si el externismo semántico resultara verdadero, se seguiría que es posible que un sujeto S no supiera que dos de sus pensamientos tienen contenidos distintos y, así, parecería en principio posible que S estuviera pensando que *p*, que fuera una alternativa relevante que estuviera pensando que *q*, y que S no fuera capaz de distinguir entre el escenario actual y el escenario alternativo relevante. Y, según el argumento de la discriminación, eso es suficiente para concluir que S no sabe qué está pensando.

El argumento se sustenta básicamente en dos ideas: que el externista está comprometido a negar que el contenido es transparente y que el conocimiento requiere ciertas capacidades discriminatorias:

> (Discp) [58]: para toda proposición *p*, si hay un escenario relevante W' en el que *q* es verdadero y *p* es falso, tal que S no puede distinguir entre el escenario W' y el escenario actual W (en el cual *p* es verdadero), entonces S no sabe que *p*.

[57] According to the discrimination argument, anti-individualism threatens privileged access by undermining a subject's ability to distinguish a priori between the thought contents she actually has and the thought contents she would have in various counterfactual situations. (Brown (2004), p. 37)

[58] La formulación es nuestra; Brown en ningún momento formula explícitamente el principio sobre conocimiento y discriminación que acepta (no creemos que el principio (Discp) presentado por nosotros se aleje mucho de la versión concreta del principio que acepte Brown).

Pongamos un ejemplo. Supongamos que Coraline ve a su madre de espaldas al final del pasillo—sobre esa percepción, llega a la creencia de que su madre está al final del pasillo. Ahora, es una alternativa relevante que no sea la madre de Coraline quien está al final del pasillo, sino Su Otra Madre, un ser pálido, muy parecido a su madre (excepto en que tiene botones negros en vez de ojos), que habita la casa en la que vive Coraline. Dado que Coraline no puede distinguir entre su madre y Su Otra Madre cuando éstas se encuentran de espaldas, no puede distinguir entre el escenario actual y el escenario alternativo relevante en el que es Su Otra Madre la que se encuentra en el pasillo. Por eso, sobre la base de (Discp), se sigue que Coraline no sabe que su madre está al final del pasillo (aunque la ve).

Por otro lado, ya hemos mencionado que sí parece haber al menos cierta tensión entre el externismo semántico y la transparencia del contenido. Brown caracteriza de la siguiente manera las dos tesis sobre la transparencia del contenido (ya en la introducción hemos mencionado esta versión concreta):

> **Transparencia de mismidad de contenido:** para cualesquiera dos pensamientos, o constituyentes de pensamiento, que *S* considera en el momento *t*, si tienen el mismo contenido, entonces, en *t*, *S* puede darse cuenta a priori de que tienen el mismo contenido.
> **Transparencia de diferencia de contenido:** para cualesquiera dos pensamientos, o constituyentes de pensamiento, que *S* considera en el momento *t*, si tienen contenidos diferentes, entonces, en *t*, *S* puede darse cuenta a priori de que tienen contenidos diferentes.

Si es verdad que el externista está comprometido a negar la tesis de la transparencia de diferencia de contenido, parece en principio posible que un sujeto S esté pensando que *p*, y que haya un escenario relevante W' en el que S no está pensando que *p*, tal que S no puede darse cuenta, sin ayuda de ninguna evidencia empírica, de que el escenario W' y el escenario actual W son diferentes. Si se dieran tales condiciones entonces, sobre la base de (Discp), S no podría saber mediante introspección que está pensando que *p*.

Volvamos con Oscar. ¿Puede éste saber mediante introspección que está pensando que el agua se hiela a cero grados? Dado su historial de transiciones entre la Tierra y la Tierra Gemela, el que esté en la Tierra Gemela pensando que la bi-agua se hiela a cero grados parece un escenario relevante, y parece bastante plausible que no pueda distinguir entre el escenario actual y el alternativo. Por ello, sobre la base de (Discp), el

incompatibilista concluye que Oscar no puede saber, mediante introspección, que está pensando que el agua se hiela a cero grados. El externismo semántico y el auto-conocimiento autoritativo parecen de nuevo incompatibles.

Mencionemos un último punto, para terminar de presentar el argumento. Boghossian condicionaba la incompatibilidad entre el externismo semántico y el auto-conocimiento autoritativo a la aceptación del modelo observacional del auto-conocimiento. Brown no dice nada acerca de estas cuestiones. Pero nos gustaría recalcar que, a pesar de ello, sí parece que el argumento presupone al menos ciertos paralelismos entre el auto-conocimiento y el conocimiento perceptivo. Parece razonable que principios del estilo de (Discp) se impongan al conocimiento perceptivo; si dos escenarios son indistinguibles para S, parece que S obtendría la misma evidencia perceptiva en los dos escenarios y, por eso, que S pudiera conocer sobre la base de esa evidencia que se encuentra en alguno de los dos escenarios parecería arbitrario. El argumento de la discriminación presupone que un mismo principio impone condiciones tanto al auto-conocimiento como al conocimiento perceptivo; sí presupone un modelo (al menos en cierta medida) observacional del auto-conocimiento.

2.2. (RA) y (DISCP)

Ése era, pues, el argumento de la discriminación. Cuando Brown introduce el argumento, no resulta claro si piensa que está presentando el mismo argumento que Boghossian, una versión distinta de ese mismo argumento, u otro argumento distinto. En esta sección defenderemos que hay una estrecha relación entre (RA) y (Discp), y que esto al menos sugiere que los dos argumentos no difieren demasiado entre sí.

(Discp) viene a decir que si un sujeto no puede *discriminar* (o *distinguir*) entre su escenario actual W y un escenario alternativo relevante W', entonces no puede saber que *p* (donde *p* es verdadero en W pero no en W'). Ahora, ¿qué quiere decir que un sujeto *discrimine* entre dos escenarios distintos? Brown no dice nada al respecto. Creemos que Williamson (1990) ofrece una caracterización adecuada:

> *a* es indiscriminable de *b* para un sujeto en un momento si y sólo si en ese momento el sujeto no es capaz de discriminar entre *a* y *b*, esto es, si y sólo si en ese momento el sujeto no es capaz de activar (adquirir o emplear) la clase de conocimiento relevante de que *a* y *b* son distintos.[59]

Lo dicho: discriminar entre *a* y *b* consiste en "activar" el conocimiento de que *a* y *b* son dos objetos distintos, en adquirir ese conocimiento que no tenía uno o, si ya lo tenía, emplearlo, "traerlo a la mente".

Además, de nuevo con Williamson (1990, 2000), creemos que uno discrimina entre escenarios (u objetos, o proposiciones) *bajo modos de presentación*: uno podría ser capaz de discriminar entre *a* y *b* si se le presentaran bajo los modos *x* e *y*, y no poder hacerlo así si se le presentaran bajo los modos *z* y *k*.

Pongamos ejemplos. Supongamos que Ahab ve a Queequeg a una distancia de cinco metros. El modo bajo el cual se le presenta (la percepción que tiene de él) es tal que Ahab puede activar su conocimiento de que ese objeto, Queequeg, no es Tashtego (cuando se le presenta bajo el modo "El arponero Tashtego"). Pero supongamos que Ahab ve a Queequeg a una distancia de doscientos metros. En este caso, el modo bajo el cual se le presenta (la percepción que tiene de él) es tal que no puede activar el conocimiento de que ese objeto, Queequeg, no es Tashtego (para cualquier modo de presentación bajo el cual se le puede presentar Tashtego). Ahora, supongamos que le decimos a Ahab que uno de sus arponeros no durmió anoche (Queequeg, aunque Ahab no lo sabe), y que uno de sus arponeros casi se ahoga (Tashtego, aunque Ahab no lo sabe), ¿puede en este caso discriminar Ahab entre Queequeg y Tashtego? Parece que no, tal y como se le presentan estos objetos, Ahab no puede activar su conocimiento de que son distintos (el arponero que no durmió puede ser el arponero que casi se ahoga).

Resumiendo, uno discrimina entre los objetos *a* y *b*, cuando se le presentan bajo los modos *x* e *y*, si y sólo si activa el conocimiento de la proposición de que *x/a* e *y/b* son dos objetos distintos. La cuestión es que esta interpretación de lo que significa que uno discrimine entre dos objetos (la más plausible que conocemos) está en consonancia con una estrecha relación entre evidencia y discriminación: uno puede activar el

[59] *a* is indiscriminable from *b* for a subject at a time if and only if at that time the subject is not able to discriminate between *a* and *b*, that is, if and only if at that time the subject is not able to activate (acquire or employ) the relevant kind of knowledge that *a* and *b* are distinct. (Williamson (1990), p. 8)

conocimiento de la proposición de que *x/a* e *y/b* son dos objetos distintos sólo si la evidencia que tiene basta para justificar su creencia de que esos dos objetos (bajo sus modos de presentación correspondientes) son distintos, sólo si la evidencia que tiene es incompatible con que esos dos objetos (bajo los modos de presentación correspondientes) sean el mismo.

Al menos así parece seguirse de los ejemplos expuestos en el párrafo anterior. En el primer ejemplo Ahab podía discriminar entre Queequeg y Tashtego porque podía activar su conocimiento de que el hombre que ve a cinco metros no es el arponero Tashtego. Y, creemos, la respuesta más simple a cómo es posible que Ahab active este conocimiento es que la evidencia perceptiva que tiene es incompatible con que sea Tashtego el hombre que tiene a cinco metros. Por otro lado, en el último ejemplo hemos dicho que Ahab no puede discriminar entre Queequeg y Tashtego porque no puede activar el conocimiento de que el arponero que no durmió es el arponero que casi se ahoga. De nuevo, creemos que la respuesta más sencilla a por qué Ahab no puede activar este conocimiento es que la evidencia que tiene (el testimonio de que uno de sus arponeros no durmió y que uno de sus arponeros casi se ahoga) es compatible con que el arponero que no durmió sea el mismo que casi se ahoga. Hay una estrecha relación entre la evidencia que tiene uno y las capacidades discriminatorias de las que dispone.

Volvamos al ejemplo de Oscar, y expliquemos por qué según el incompatibilista se sigue, sobre la base de (Discp), que Oscar no sabe qué está pensando. Con lo dicho, sabemos que algo más se debería decir sobre en qué medida Oscar no discrimina entre la Tierra y la Tierra Gemela, ya que sí hay dos modos de presentación (las descripciones "La Tierra, donde hay agua y no bi-agua, y donde estás ahora" y "La Tierra Gemela, donde hay bi-agua y no agua, y donde has estado antes") tal que, si a Oscar se le presentaran la Tierra y la Tierra Gemela bajo esas descripciones, podría descubrir que la Tierra y la Tierra Gemela son distintas.

Por supuesto, el argumento de la discriminación no dice que Oscar no puede llegar a conocer que está pensando que el agua se hiela a cero grados, dice que no puede saberlo *mediante introspección*, porque es incapaz de discriminar entre su escenario actual y el escenario alternativo relevante *mediante introspección*. Esto nos da alguna pista sobre cuáles son los modos de presentación relevantes para Oscar: en el escenario alternativo

Oscar estaría en el mismo estado fenoménico en el que de hecho se encuentra, ese estado fenoménico, parece, constituye un modo de presentación que comparten los dos escenarios, y cuando se le presentan bajo ese modo de presentación Oscar no puede distinguir entre los dos escenarios (porque de hecho comparten modo de presentación).

No es evidente que del hecho de que en W y W' S esté en el mismo estado fenoménico se siga que esos dos escenarios son indistinguibles para S mediante introspección. Por eso, creemos que esta tesis constituye una premisa implícita en el argumento de Brown. Más adelante, en las secciones 5.3. y 5.4. diremos más acerca de la plausibilidad o no de esta premisa implícita, lo que nos interesa aquí es señalar cierta analogía entre el argumento de Boghossian y el de Brown. Porque el argumento de la discriminación presupone que si en W y W' S está en el mismo estado fenoménico, entonces S no puede discriminar mediante introspección entre W y W', y el argumento de Boghossian presupone que si S está en el mismo estado fenoménico en W y W', entonces S obtiene en los dos escenarios la misma evidencia introspectiva[60].

Ahora, si son ciertas las disquisiciones sobre discriminación y evidencia que se han presentado en los párrafos anteriores, se sigue que uno puede discriminar entre *a* y *b* (bajo los modos de presentación relevantes) si y sólo si la evidencia que tiene uno cuando se le presenta *a* es incompatible con que esté en un escenario donde es *b* el objeto que se le está presentando: los antecedentes de (RA) y (Discp) son equivalentes. La evidencia de uno en W será incompatible con el escenario relevante W' si y sólo si esa evidencia le basta para activar el conocimiento de que W y W' son dos escenarios distintos. Pero, dado que comparten consecuentes, se sigue que (RA) y (Discp) son dos principios epistémicos equivalentes.

Por supuesto, alguien podría rechazar las ideas presentadas, y así defender que (RA) y (Discp) no son equivalentes. No entraremos en la discusión. Nos basta con señalar que hay motivos para pensar que hay una relación muy estrecha (nosotros diríamos que equivalencia) entre (RA) y (Discp).

[60] La premisa (4) en nuestra formalización del argumento de Boghossian recoge esta idea. ((4) La evidencia que obtiene Oscar mediante introspección es la misma en el escenario actual y el escenario relevante, ya que si estuviera en la Tierra Gemela estaría en el mismo estado interno que el actual.)

2.3. ALGUNAS POSIBLES RESPUESTAS (Y CRÍTICAS DE BROWN)

2.3.1. Discriminación *a priori*.

Brown presenta como distintas dos respuestas alternativas que, en nuestra opinión, no difieren demasiado entre sí (ya que las dos vienen a decir que Oscar tiene las suficientes capacidades discriminatorias). Por eso, presentaremos ambas en esta misma subsección. La primera de ellas vendría a proponer el lema "la fiabilidad aporta discriminabilidad". Es una actitud bastante común entre los compatibilistas, dice Brown, el intentar responder al argumento haciendo hincapié en la fiabilidad de nuestros juicios sobre nuestros propios estados mentales. Así, vienen a decir, dado que son fiables, estos juicios están suficientemente justificados y constituyen conocimiento. El problema es que el argumento de la discriminación en ningún momento protestaba que nuestros juicios sobre nuestros estados mentales no son fiables (Brown acepta que lo son); el compatibilista estaría ofreciendo fiabilidad cuando se le ha pedido discriminabilidad.

Esta primera respuesta viene a decir que la fiabilidad *aporta* discriminabilidad, esto es, que si la creencia de S de que *p* se basa en un método de formación de creencias fiable, entonces S tiene las capacidades discriminatorias que se le podrían exigir. De hecho, parece que algunos autores fiabilistas (Goldman (1976)) defienden que la fiabilidad y la discriminabilidad "vienen la una de la mano de la otra", que la creencia de que *p* de uno es fiable si y sólo si puede distinguir entre el escenario actual en el que *p* es verdadero y todo escenario relevante en el que es falso. Por ejemplo, parece que esto es así en el caso del conocimiento empírico basado en percepción; parece que la creencia de uno de que lo que tiene delante es una manzana será fiable si y sólo si puede distinguir entre manzanas y peras (si es un escenario relevante que tenga una pera delante), o si y sólo si puede distinguir entre una manzana de verdad y una de plástico (si es un escenario relevante que lo que tiene delante sea una manzana de plástico).

Ahora, si esto fuera así, si fuera verdad que "la fiabilidad aporta discriminabilidad", sería fácil responder al argumento de la discriminación una vez aceptamos que nuestros

juicios de auto-conocimiento son fiables, ya que se seguiría que sí tenemos esas capacidades discriminatorias que nos exige (Discp).

La segunda estrategia vendría a decir, simplemente, que Oscar sí tiene las capacidades discriminatorias que se le exigen (independientemente de que sus juicios de auto-conocimiento sean fiables o no).

Pero Brown rechaza que Oscar tenga tales capacidades discriminatorias, ya que carece de las "habilidades asociadas con la posesión de ciertas capacidades discriminatorias—la habilidad de notar los cambios y la habilidad de hacer juicios de mismo/diferente. Esto puede crear dudas en la tesis de que puede distinguir a priori entre los dos tipos de pensamiento"[61]. Y, sobre la tesis de que "la fiabilidad aporta discriminabilidad", sucede que nos es bastante fácil encontrar contraejemplos a la ecuación propuesta:

> El caso de transición lenta nos muestra que la capacidad fiable para formar creencias verdaderas acerca de cuál de dos tipos es instanciada no es suficiente para la capacidad de distinguir entre esos dos tipos. [62]

Coincidimos con Brown cuando niega que Oscar tiene las capacidades discriminatorias que algunos le adscriben. Ahora, también creemos que la discusión de Brown sobre el tema ignora algunas cuestiones importantes—por ejemplo, no discute si estar en relación de *acquaintance* con un objeto basta para estar en posición de discriminar entre ese objeto y cualquier otro (Evans (1982)), ni en qué medida se podría aceptar que Oscar está en esta relación de *acquaintance* con su pensamiento. Trataremos estas cuestiones más adelante (secciones 5.2. y 5.3.).

2.3.2. La fiabilidad basta ((Discp) es falso)

Otra posible respuesta viene a defender que la fiabilidad basta para el conocimiento, y que los principios del tipo (Discp) son falsos. El compatibilista que siguiera esta estrategia aceptaría que alguien en la situación de Oscar no podría distinguir sin ayuda de investigación empírica entre los dos escenarios relevantes, pero negaría que esto

[61] ...abilities associated with possession of a discriminative capacity—the ability to notice change and the ability to make correct same/different judgments. This may cast doubt on the claim that she can distinguish a priori the two types of thought. (Brown (2004), p. 52)
[62] ...the slow switch case itself shows us that the reliable ability to form true beliefs about which of two types is instantiated is not sufficient for the ability to distinguish those two types. (Brown (2004), p. 48)

fuera una amenaza para su auto-conocimiento autoritativo. Sabe qué está pensando porque su creencia es fiable, y el hecho de que no pueda distinguir entre el escenario actual y el escenario alternativo no mina su conocimiento, ya que (Discp) es falso.

Brown no da demasiados argumentos en contra de esta estrategia; simplemente se limita a señalar que el compatibilista no podría basar su defensa en los ejemplos fiabilistas clásicos, ejemplos de conocimiento empírico. Defiende que esos ejemplos igualmente podrían apoyar una posición "discriminabilista" de la justificación y el conocimiento, ya que en los ejemplos parece claro que si el sujeto no tiene una evidencia fiable, eso es así porque carece de ciertas capacidades discriminatorias ("en estos casos, el hecho de que un sujeto tenga un proceso de formación de creencias fiable puede explicarse por el hecho de que el sujeto puede distinguir entre las clases relevantes"[63]).

Creemos que la crítica de Brown no es satisfactoria. Dejando a un lado las cuestiones concernientes a la fiabilidad, el compatibilista podría intentar proporcionar más argumentos contra (Discp); concluir que una posición como la esbozada no resulta satisfactoria por el mero hecho de que no se podría basar en los ejemplos fiabilistas clásicos resulta, al menos, precipitado. En el capítulo sobre Falvey y Owens (1994) intentaremos defender que tanto (Discp) como (RA) son falsos (sección 4.2.).

2.3.3. Reemplazo de conceptos

Hay dos modos distintos de interpretar las transiciones lentas. El primero[64] afirma que cuando Oscar viaja a la Tierra Gemela y, con el tiempo, adquiere el concepto BI-AGUA, no pierde su concepto anterior AGUA—esto es, según esta opción, los casos de transición lenta son casos de *cohabitación de conceptos*. La segunda interpretación[65] defiende que las transiciones lentas son ejemplos de *reemplazo conceptual*—Oscar pierde su concepto AGUA cuando adquiere el concepto BI-AGUA, y no hay ningún momento en la historia de Oscar (según la hemos contado) en el que éste tenga los dos conceptos[66]. El

[63] in these cases, the fact that a subject has a reliable belief-forming process can be explained by the fact that the subject can distinguish the relevant kinds. (Brown (2004), p. 63)
[64] Burge (1998), Gibbons (1996), Schiffer (1992), Boghossian (1992a, 1994)
[65] Ludlow (1995b, 1996, 1999), Falvey y Owens (1994), Tye (1998), Bernecker (1998), Brueckner (1997)
[66] En la segunda parte, en la que en gran parte nos centraremos en cuestiones acerca de las consecuencias del externismo semántico sobre la memoria, trataremos estas cuestiones de cohabitación y reemplazo con más detenimiento (Véanse, especialmente, las secciones 2.1. y 2.2. y el capítulo 6).

incompatibilista podría intentar sugerir una respuesta al argumento de la discriminación basándose en la interpretación según la cual los casos de transición lenta son casos de reemplazo conceptual:

> La aparente incapacidad [de Oscar para discriminar] podría explicarse, no aduciendo que carece de esta capacidad, sino acudiendo a cierta incapacidad, después de la transición, para recordar los pensamientos que solía expresar con 'agua'. (...) Podría sugerirse que el motivo por el cual Oscar no puede hacer juicios correctos de mismidad o diferencia entre pensamientos de agua y bi-agua es que, después de la transición, ya no puede recordar los pensamientos que solía expresar con 'agua'.[67]

Esto es, no es que Oscar no tenga las capacidades discriminatorias relevantes, sino que no es capaz de "traer a la mente" el pensamiento de bi-agua en cuestión. Brown niega de nuevo que Oscar pueda distinguir entre sus pensamientos de agua y bi-agua, porque "daría la misma explicación de sus conceptos de agua y bi-agua; diría de cada uno de ellos que es el concepto del líquido inodoro e incoloro típico en lagos y ríos"[68] y, por lo tanto, "el modo en el que [Oscar] describe sus pensamientos proporciona evidencia a favor de que no puede distinguir a priori entre pensamientos de agua y bi-agua"[69].

El defensor del reemplazo podría proponer una respuesta más convincente que la que caracteriza Brown, por eso dedicaremos algunos párrafos a describir esta opción que, aunque a primera vista pueda parecer atractiva, creemos que tiene problemas[70]. El defensor del reemplazo podría defender que, dado que carece de las herramientas conceptuales necesarias para poder siquiera plantearse el escenario alternativo, no ha lugar para exigirle a Oscar que distinga entre el escenario actual y el alternativo o para exigirle que excluya de algún modo ese escenario alternativo. Esto es, según esta estrategia, dado que las transiciones lentas son ejemplos de reemplazo, los escenarios alternativos que podrían minar el auto-conocimiento del sujeto no son *relevantes*.[71]

---

[67] [Oscar's] apparent inability [to discriminate] might be explained not by [his] lacking this ability, but rather by an inability, post-switch, to remember the thoughts [he] used to express with 'water'. (...) It may be suggested that the reason [Oscar] cannot make correct judgments of sameness or difference between water and twater thoughts is that, after the switch, [he] can no longer remember the thoughts [he] used to express with 'water'. (Brown (2004), p. 55)

[68] ...would give the same explanations of her concepts of water and twater; [he] would say of each that it is the concept of the odorless, colorless liquid common to lakes and rivers. (Brown (2004), p. 57)

[69] ...the way [Oscar] describes [his] thoughts provides evidence that [he] cannot distinguish a priori between water and twater thoughts. (Brown (2004), p. 57)

[70] Pérez Otero (2009) propone una posición que, aunque no es exactamente la que presentamos nosotros en los párrafos siguientes, se le asemeja bastante.

[71] Parece que Boghossian tenía algo así en mente cuando escribió lo siguiente: "No constituye una objeción al argumento señalar que, en *este* modo de relatar el ejemplo de la transición, S ni siquiera puede

Creemos que esta versión de la respuesta basada en el reemplazo nos lleva a algunos resultados bastante satisfactorios. Pero no está exento de problemas. Primero, tengamos en mente que el quid de esta respuesta viene a ser que da igual cuánto viaje Oscar entre la Tierra y la Tierra Gemela, los escenarios alternativos no serán relevantes. El problema es que no está claro qué podría motivar esta idea. Uno no puede decidir que alguien como Oscar perdería el concepto anterior al adquirir el nuevo y que por eso el escenario anterior ya no es relevante, al contrario, parece que los viajes de Oscar hacen que los escenarios alternativos sean relevantes y, por eso, la idea de que Oscar reemplaza el concepto anterior por uno nuevo pierde fuerza. Por ejemplo, parece claro que si preguntáramos a Oscar acerca de ejemplares de "agua" que forman el estándar de su concepto (BI-)AGUA, éste mencionaría tanto ejemplares actuales de agua como ejemplares pasados de bi-agua. Heal (1998) desarrolla esta idea, concluyendo que en pocas ocasiones se da completamente una transición lenta:

> Bajo qué circunstancias es, por lo tanto, una transición completa? (...) se completará cuando ninguno de los especímenes citados sea agua. Pero eso ocurrirá sólo cuando la víctima de la transición haya perdido el contacto cognitivo con la Tierra, esto es, ha perdido toda capacidad para pensar, bajo cualquier tipo de descripción identificadora, sobre especímenes que de hecho son agua y/o ha perdido toda efectividad en etiquetar esos especímenes como 'agua'. La consecuencia de esto es que no tenemos razones para suponer que alguna vez se garantiza que una transición se completa.[72]

---

plantearse la hipótesis que se le está pidiendo que excluya. Uno podría no tener el concepto de moneda falsa, pero si hay muchas monedas falsas en su entorno, entonces tiene que ser capaz de excluir la hipótesis de que la moneda en su mano es falsa para que podamos decir que sabe que es un penique. El hecho de que no pueda plantearse la hipótesis relevante no le absuelve de este requerimiento." ("It is no objection to this argument to point out that, on *this* way of telling the switching story, S cannot even frame the hypothesis he is called upon to exclude. Someone may not have the concept of counterfeit money, but if there is a lot of counterfeit money in his vicinity, then he must be able to exclude the hypothesis that the coin in his hand is counterfeit before he can be said to know that it is a dime. The fact that he cannot so much as frame the relevant hypothesis does not absolve him of this requirement." (Boghossian (1989), p. 160, n. 12)). Pero la estrategia basada en reemplazo que tenemos en mente tiene una buena respuesta a mano porque, como dice Pérez Otero (2009): "Por no tener el concepto alternativo requerido, *necesario para representarse canónicamente el contenido proposicional correspondiente*, no hay riesgo de error cuando Oscar considera su pensamiento. En el caso de la moneda falsa, lo importante no es tener o no el concepto de moneda falsa, sino que el sujeto tendrá un medio de representarse (perceptivamente) la moneda falsa, con el consiguiente riesgo de error." (Pérez Otero (2009), p. 231). Esto es, la cuestión no es simplemente que Oscar carece del concepto BI-AGUA, sino que, al carecer de ese concepto, no es factible que Oscar se represente un pensamiento de bi-agua. Y esto introduce una diferencia importante entre la situación de Oscar y la situación de quien carece del concepto MONEDA FALSA en el ejemplo de Boghossian: este último puede tener una percepción de una moneda falsa aun y cuando no tiene el concepto MONEDA FALSA, Oscar en cambio no puede tener un pensamiento de bi-agua. Dado que, al no tener las herramientas conceptuales necesarias para ello, no es factible que Oscar esté pensando que la bi-agua se hiela a cero grados, ése es un escenario que, con el bagaje conceptual actual de Oscar, no se puede dar. Da igual cuánto haya viajado Oscar: si no tiene el concepto BI-AGUA, el escenario alternativo no puede ser relevante.

[72] Under what conditions, then, is a switch complete? (...) it will be complete when none of the specimens cited is water. But that will occur only when the victim of the switch has lost cognitive contact with

Coincidimos en parte con Heal (1998), ya que como ella opinamos que si las transiciones lentas fueran ejemplos de reemplazo conceptual, entonces sería difícil que una transición se diera *completamente*[73]. Pero, a diferencia de ella, creemos que esto motiva la interpretación según la cual las transiciones lentas son ejemplos de cohabitación conceptual; Oscar forma parte de una nueva comunidad lingüística, y nos gustaría decir que adquiere el nuevo concepto que emplean en esa comunidad[74]—si resulta problemático mantener que perdió su antiguo concepto porque de hecho no ha perdido todo el contacto cognitivo que tenía con su entorno anterior, la conclusión más simple se reduce a asumir que en las transiciones hay cohabitación de conceptos.

Pero no es ése el único problema que creemos tiene el reemplazo conceptual y, por eso, esta respuesta al argumento nos parece poco plausible. Pero no entraremos a enumerar los problemas del reemplazo aquí, dejaremos esta tarea para la segunda parte del trabajo (secciones 6.2.2. y 6.2.3.). Mencionemos, simplemente, que resulta curioso que esta estrategia basada en el reemplazo que niega que los escenarios alternativos sean relevantes se asemeja bastante a la respuesta al argumento que al final propone Brown. Presentaremos esta respuesta más adelante, en el capítulo 6, así que no diremos nada más aquí. Simplemente, mencionemos que creemos que tanto Brown como la respuesta esbozada en esta sección tienen un problema en común, relacionado con los pensamientos basados en ostensión—elaboraremos esta crítica más adelante (sección 6.4.1.).

Earth, i.e. has lost all ability to think, under any identifying description whatsoever, of specimens which are in fact water and/or has lost all confidence that those specimens deserve the label 'water'. The implication of this is that we have no reason to suppose that a switch is guaranteed ever to be complete. (Heal (1998), p. 107)

[73] Bueno, no del todo. Uno puede defender que, en parte debido a estas cuestiones, en una transición lenta uno reemplaza su antiguo concepto por un *concepto amalgama* (véase la sección 2.2 de la segunda parte).

[74] Heal (1998) explícitamente dice que trata sólo cuestiones relacionadas con el externismo de clases naturales, y por eso deja de lado factores como qué comunidad lingüística rodea a Oscar. Creemos que éste es el motivo principal por el cual no llega a apostar por un modelo de cohabitación.

3. TYLER BURGE

Burge defiende una posición compatibilista que presenta principalmente en "Individualism and Self-Knowledge", de 1988—el artículo es, pues, anterior a las dos versiones del argumento incompatibilista que hemos presentado (Boghossian (1989) y Brown (2004)). Es en "Individualism and Self-Knowledge" donde por primera vez alguien presenta un escenario de transición lenta y es, por lo tanto, también la primera vez que se ofrece una respuesta para los posibles argumentos incompatibilistas que se basan en tales ejemplos. Burge caracteriza lo que denomina 'pensamientos *cogito*', y defiende que, en cuanto son auto-verificantes, ningún argumento incompatibilista podrá minar el conocimiento sobre nuestros estados mentales que obtenemos mediante esos pensamientos.

3.1. PENSAMIENTOS *COGITO* Y AUTO-CONOCIMIENTO

Burge (1988) no propone un ejemplo concreto de transición lenta, pero sí que menciona en qué medida un escenario alternativo podría minar el auto-conocimiento de un individuo que viaja de una Tierra a la otra constantemente, pasando suficiente tiempo

cada vez que transita a un nuevo escenario. Acepta que estos ejemplos nos llevan a cierta "sensación de que hay un rompecabezas ahí", que uno "siente" que, al menos en esos casos de transición lenta, el externismo semántico y el auto-conocimiento autoritativo podrían ser incompatibles, y además parece que tiene en mente una versión del argumento incompatibilista cercana a la de Brown (2004)—no dice explícitamente cuál es el argumento que pretende responder, pero sí que dice que las incapacidades discriminatorias de la víctima de la transición lenta nos llevan a pensar erróneamente que no puede conocer mediante introspección qué está pensando:

> [La víctima de una transición lenta] no tendría ninguna señal de las diferencias en sus pensamientos, ninguna diferencia en el modo en el que las cosas "se sienten". (...) El resultado de todo esto es que el individuo tendría pensamientos diferentes dependiendo de las transiciones, pero no sería capaz de comparar las situaciones y conocer cuándo y dónde ocurrían las diferencias. Esto sugiere fácilmente, aunque también creo que equivocadamente, la posterior tesis de que tal individuo no podría conocer qué pensamientos tiene a menos que emprendiera una investigación empírica del entorno que sacara a la luz las diferencias entre los entornos.[75]

De acuerdo con Burge, los ejemplos de transición lenta son importantes porque caracterizan a individuos que, en dos escenarios relevantes distintos, tienen pensamientos distintos a pesar de que los dos escenarios "se sienten" igual, son fenoménicamente indistinguibles. Alguien podría defender que este "sentirse igual" de los dos escenarios mina el auto-conocimiento autoritativo del individuo en cuestión (así lo hacen Boghossian (1989) y Brown (2004)), aunque esta conclusión sería demasiado precipitada de acuerdo con Burge—la víctima de la transición lenta podría tener aquello que etiqueta como "pensamiento *cogito*". Los pensamientos *cogito* son pensamientos de segundo orden, cuya peculiaridad es que están en parte constituidos por el pensamiento de primer orden acerca del cual son:

> [El paradigma de pensamientos *cogito* incluye] no sólo 'estoy pensando ahora', sino también 'pienso (con este mismo pensamiento) que escribir requiere concentración' y 'juzgo (o dudo) que el agua es más común que el mercurio'. (...) Resulta realmente plausible que este tipo de juicios o pensamientos constituyen conocimiento, que no son producto de investigación empírica ordinaria, y que son peculiarmente directos y autoritativos. Es más, los juicios de este tipo son auto-

[75] [The victim of a slow switch] would have no signs of the differences in his thoughts, no difference in the way things "feel". (...) The upshot of all this is that the person would have different thoughts under the switches, but the person would not be able to compare the situations and note when and where the differences occurred. This point easily, though I think mistakenly, suggests the further point that such a person could not know what thoughts he had unless he undertook an empirical investigation of the environment which would bring out the environmental differences. (Burge (1988), p. 653)

> verificantes en un modo obvio: el mero hacer estos juicios los hace verdaderos. (...) llamaré a tales juicios *auto-conocimiento básico*.[76]

Los pensamientos *cogito* son pensamientos de segundo orden que incluyen el pensamiento de primer orden acerca del cual son, como por ejemplo ESTOY PENSANDO, CON ESTE MISMO PENSAMIENTO, QUE EL AGUA SE HIELA A CERO GRADOS. Dada esta relación de constitución, se sigue que no puede haber ningún tipo de discordancia entre los contenidos del pensamiento de primer orden y de todo el pensamiento *cogito*: si el pensamiento de primer orden tuviera un contenido distinto, así lo tendría también el pensamiento *cogito* entero. Por ejemplo, si en vez de estar en la Tierra pensando que el agua se hiela a cero grados, Oscar estuviera en la Tierra Gemela pensando que la bi-agua se hiela a cero grados, entonces Oscar no tendría el pensamiento *cogito* ESTOY PENSANDO, CON ESTE MISMO PENSAMIENTO, QUE EL AGUA SE HIELA A CERO GRADOS, sino el pensamiento *cogito* ESTOY PENSANDO, CON ESTE MISMO PENSAMIENTO, QUE LA BI-AGUA SE HIELA A CERO GRADOS. Así, debido a que no hay posibilidad para una discordancia entre los contenidos, los pensamientos *cogito* son auto-verificantes: el mero pensarlos los hace verdaderos. Uno puede ver esto *a priori* y, por eso, sabe que pensar un pensamiento *cogito* asegura que está pensando el pensamiento de primer orden que en parte lo constituye: este tipo de pensamientos proporciona conocimiento.

Los pensamientos *cogito* están en parte constituidos por el pensamiento de primer orden acerca del cual son; por lo tanto, presuponen al menos las mismas condiciones necesarias para que se dé el pensamiento de primer orden—en cuanto el pensamiento de primer orden y el pensamiento *cogito* se dan, es necesario que las condiciones necesarias para que se den esos dos pensamientos también se den. El mismo darse del pensamiento *cogito* presupone que se dan las condiciones necesarias para que se dé el pensamiento de primer orden y, por eso, para saber que está pensando que *p*, el individuo que tenga un pensamiento *cogito* sobre el pensamiento de que *p* no necesita saber primero que de hecho se dan las condiciones necesarias para que piense que *p*.

---

[76] [The paradigm of *cogito* thoughts includes] not merely 'I am now thinking', but 'I think (with this very thought) that writing requires concentration' and 'I judge (or doubt) that water is more common than mercury'. (...) It is certainly plausible that these sorts of judgments or thoughts constitute knowledge, that they are not products of ordinary empirical investigation, and that they are peculiarly direct and authoritative. Indeed, these sorts of judgments are self-verifying in an obvious way: making these judgments itself makes them true. (...) I shall call such judgments *basic self-knowledge*. (Burge (1988), p. 649)

El último párrafo parecía un trabalenguas; pongamos un ejemplo para hacerlo más claro. Supongamos que Coraline piensa el pensamiento *cogito*: ESTOY PENSANDO, CON ESTE MISMO PENSAMIENTO, QUE LAS ARAÑAS SON FEAS. Ese pensamiento está en parte constituido por el pensamiento de primer orden LAS ARAÑAS SON FEAS. Supongamos que para que Coraline pueda pensar ese pensamiento de primer orden es necesario que se dé la condición C (un buen candidato es que las arañas de hecho existan). ¿Debe Coraline saber primero que C es una condición necesaria del pensamiento de primer orden y que de hecho se da la condición C, para luego saber qué está pensando? No. Dado que el pensamiento *cogito* está en parte constituido por ese pensamiento de primer orden, la condición necesaria C también será una condición necesaria del pensamiento *cogito* (si no se diera C, Coraline no pensaría ese pensamiento *cogito*, sino otro); dado que Coraline sí está pensando ese pensamiento *cogito*, puede saber que se da toda condición K necesaria para que piense LAS ARAÑAS SON FEAS (si no se diera, no estaría pensando el pensamiento *cogito* en cuestión). Por eso, no necesita saber que C es una condición necesaria o que de hecho se da C, sabe que cualquier condición K de hecho se da (y C es una de esas condiciones K).

Ya hemos señalado antes que todo principio defendiendo que, para saber que *p*, uno debe conocer primero todas las condiciones necesarias para que se dé que *p*, es falso[77]. También Boghossian (1989) advierte que los principios de este estilo son falsos, y asegura que su argumento no se basa en ningún principio de este tipo—por lo tanto, si el objetivo de la estrategia basada en los pensamientos *cogito* fuera que éstos demuestran que tales principios son falsos, de poco serviría. A pesar de ello, los pensamientos *cogito* no demuestran sólo que estos principios son falsos, también demuestran que ningún escenario alternativo puede minar el auto-conocimiento que se obtiene mediante los pensamiento *cogito*. Como hemos señalado, los pensamientos *cogito* son auto-verificantes—no pueden ser falsos, uno puede así *blindar* el conocimiento que tiene de los contenidos de sus estados mentales. Ningún escenario alternativo relevante puede amenazar el auto-conocimiento que obtenemos mediante los pensamientos *cogito*, porque éstos no pueden ser falsos y *sabemos* que no pueden serlo y, por lo tanto, la víctima de una transición lenta no necesita de evidencia empírica para conocer qué está

[77] En el primer capítulo llamábamos a este principio (Nec), y proponíamos el ejemplo de Maider pretendiendo saber que Isabella Rossellini es una de las protagonistas de *Blue Velvet* como contraejemplo al principio.

pensando. Ninguna teoría semántica podría amenazar el conocimiento que nos aportan estos pensamientos—el externismo semántico y el auto-conocimiento autoritativo son compatibles.

Los argumentos incompatibilistas expuestos, así como la reivindicación de que al no tener ciertas capacidades discriminatorias, la víctima de una transición lenta no podría conocer por introspección qué está pensando, se basan en un error según Burge. El error no es más que no comprender que el conocimiento perceptivo y el auto-conocimiento mediante introspección guardan diferencias más que notables.

> Quisiera llamar la atención acerca de algunos puntos fundamentales en los que difieren el conocimiento perceptivo de entidades físicas y el tipo de auto-conocimiento que hemos estado caracterizando. (...) En cualquier ocasión, nuestras percepciones podrían haber sido erróneas. El ítem físico individual con el cual interactúa uno perceptivamente en cualquier momento es fundamentalmente independiente de la percepción – y concepción – de cualquier persona. La naturaleza de la entidad física podría haber sido diferente, y los estados perceptivos y otros estados mentales de uno seguir siendo los mismos. Este hecho subyace tras un hecho normativo sobre la percepción. Estamos sujetos a cierto tipo de errores posibles sobre objetos empíricos—percepciones equívocas y alucinaciones que son "brutas".[78]

> Los casos paradigmáticos de auto-conocimiento difieren del conocimiento perceptivo (...) en el caso de los juicios del tipo *cogito*, el objeto del pensamiento de uno no está relacionado contingentemente a los pensamientos que piensa uno sobre ello. Los pensamientos son auto-referenciales y auto-verificantes. Un error basado en la discordancia entre los pensamientos de uno y su objeto simplemente no es posible en estos casos. (...) Creo que, en todos estos casos de conocimiento autoritativo, los errores brutos son imposibles.[79]

Por lo tanto, en cuanto los errores "brutos" no son posibles en el auto-conocimiento básico, las exigencias que imponen principios epistémicos como (RA) o (Discp) no están justificadas. Podrían ser acertados para el conocimiento basado en percepción, pero el auto-conocimiento difiere del conocimiento perceptivo en estos aspectos.

---

[78] I want to dwell on some fundamental ways in which perceptual knowledge of physical entities differs from the sort of self-knowledge that we have been featuring. (...) On any given occasion, our perceptions could have been misperceptions. The individual physical item that one perceptually interacts with at any given time is fundamentally independent from any one person's perceptions – and conceptions. The nature of the physical entity could have been different even while one's perceptual states, and other mental states, remained the same. This fact underlies a normative fact about perception. We are subject to certain sorts of possible errors about empirical objects—misperceptions and hallucinations that are "brute". (Burge (1988), p. 657)

[79] The paradigmatic cases of self-knowledge differ from perceptual knowledge (...) in the case of cogito-like judgments, the object, or subject matter, of one's thoughts is not contingently related to the thoughts one thinks about it. The thoughts are self-referential and self-verifying. An error based on a gap between one's thoughts and the subject matter is simply not possible in these cases. (...) I think that, in all cases of authoritative knowledge, brute mistakes are impossible. (Burge (1988), p. 658)

3.2. LA CRÍTICA DE BROWN

Brown protesta que las estrategias compatibilistas basadas en enfatizar la fiabilidad de nuestros juicios acerca de nuestros propios estados mentales fallan como respuesta al argumento de la discriminación, ya lo hemos dicho antes. Enmarca a Burge (y a todo aquél que acuda a los pensamientos *cogito*) dentro de este modelo fiabilista, critica que ofrece fiabilidad cuando se le pedía discriminabilidad; independientemente de lo que puedan demostrar los pensamientos *cogito* de Burge, en ningún modo podrían suponer una respuesta al argumento de la discriminación tal y como lo expone ella—la de Burge es una no-respuesta.

> Podemos preguntarnos en qué medida constituye una respuesta al argumento de la discriminación el enfatizar la fiabilidad de las creencias del sujeto sobre sus contenidos de pensamiento. El argumento se centra en la noción de cierta capacidad discriminatoria, y no es obvio que esta noción pueda ser equivalente a la noción de cierta capacidad fiable para formar creencias verdaderas.[80]

Brown falla en su objetivo; no era intención de Burge ofrecer una respuesta compatibilista basada en una epistemología fiabilista. Es verdad que los pensamientos *cogito* son fiables en un grado máximo (en ningún caso podrían dar paso a creencias falsas), pero Burge en ningún momento afirmaba que los juicios de este tipo adquirían justificación debido a su fiabilidad[81]. Esto mismo critica Goldberg (2006) a Brown:

> Aunque seguramente una reconstrucción tal [("la reconstrucción fiabilista" de la posición de Burge)] es fiel al espíritu de varias respuestas[82] al problema, no le es

[80] We might wonder how stressing the reliability of the subject's beliefs about her thought contents constitutes a reply to the discrimination argument. That argument focuses on the notion of a discriminative ability, and it is not obvious that this notion can be equated with the notion of a reliable ability to form true beliefs. (Brown (2004), p. 44)

[81] Nos parece importante tratar estas cuestiones aquí, entre otras cosas, porque es bastante común interpretar la respuesta de Burge como una estrategia basada en una epistemología fiabilista. Bernecker (1998) por ejemplo comete este mismo error: "La teoría inclusiva del auto-conocimiento se basa en una concepción fiabilista del conocimiento. El *fiabilismo* es la teoría de que para saber algo todo lo que se requiere es que la relación creencia-hecho sea fiable. (...) Por ello, para tener auto-conocimiento un sujeto sólo necesita estar, de hecho, en alguna relación causal con sus estados de primer orden; no necesita saber que lo está." ("The inclusion theory of self-knowledge relies on a reliabilist conception of knowledge. *Reliabilism* is the view that to know something all that is required is that the belief-fact link is reliable. (...) Therefore, to have self-knowledge a subject need only, as a matter of fact, stand in some causal relation to his first-order states; he need not know that he does." (Bernecker (1998), p. 338))

[82] Gibbons (1996) es un ejemplo claro de respuesta compatibilista basada en una epistemología fiabilista; Falvey y Owens 1994) (y, quizás, también McLaughlin y Tye (1998)) también apuestan por una epistemología fiabilista, pero su respuesta compatibilista tiene también otros matices.

> fiel al espíritu de la respuesta propia de Burge. Al contrario, Burge articula su respuesta en dos pasos. Primero, los [pensamientos *cogito*] disfrutan de un tipo de apoyo epistémico que es tanto *garantía de verdad* como *accesible introspectivamente* en el siguiente sentido: un sujeto que poseyera todos los conceptos relevantes (semánticos y epistémicos) estaría en posición de apreciar "desde el sillón" la naturaleza de *garante de verdad* de los [pensamientos *cogito*]. Y, segundo, los sujetos que hacen tales juicios disfrutan de cierta legitimación epistémica[83] para ello. Dado que se apela a esta noción de legitimación epistémica, esta respuesta en dos partes al argumento incompatibilista no se caracteriza adecuadamente si se hace entendiéndolo como dependiendo en la suficiencia (para el conocimiento) de la mera fiabilidad trans-mundana.[84]

Intentemos ponerlo un poquito más simple. La respuesta de Burge se basa en una concepción *inclusiva* del auto-conocimiento, no en una concepción *fiabilista*. Esto es, cuando uno juzga mediante un pensamiento *cogito*, el pensamiento de primer orden constituye en parte el pensamiento *cogito*. Esto hace que el juicio sea "auto-verificante", no puede ser que sea falso, y el sujeto que hace el juicio puede comprender que esto es así mediante introspección. Por ello, el sujeto adquiere cierta "legitimación epistémica" para saber qué está pensando. Y la cuestión es que en ningún momento hemos apelado a la *fiabilidad* de los pensamientos *cogito*, a algún tipo de relación causal que podría haber entre ese juicio y el pensamiento de primer orden acerca del cual es; ni siquiera lo hemos hecho cuando decíamos que son auto-verificantes; el sujeto adquiere justificación (su "legitimación epistémica") mediante otras vías, no depende directamente de la fiabilidad de los pensamientos *cogito*. Por eso, entendemos que es erróneo interpretar a Burge como ofreciendo una respuesta basada en una epistemología fiabilista. Burge (1996) es explícito al respecto:

> ¿De dónde viene la legitimación [para saber qué está pensando uno]? ¿Y qué la hace capaz de sobrevivir tales cambios de entorno [p. ej.: transiciones lentas]? Opino que la legitimación relevante no deriva de la fiabilidad de alguna relación causal-perceptiva entre la cognición y su objeto. Tiene otras dos fuentes. Una es el rol de los juicios relevantes en el razonamiento crítico. La otra es la relación constitutiva entre los juicios y aquello acerca del cual son – o entre los juicios sobre los pensamientos de uno y la naturaleza veraz de los juicios.[85]

---

[83] Traducimos 'epistemic entitlement' como 'legitimación epistémica'.

[84] Although such a reconstruction is arguably in the spirit of some replies to the achievement problem, is not in the spirit of Burge's own reply. On the contrary, the thrust Burge's reply to the achievement problem is two-fold. First, first-person present-tense judgements enjoy a sort of epistemic support that is both *truth-guaranteeing* and *introspectively accessible* in the following sense: a subject who possessed all of the relevant (semantic and epistemic) concepts would be in a position to appreciate from the armchair the truth-guaranteeing nature of first-person present-tense judgments. And second, subjects who make such judgements enjoy an *epistemic entitlement* to do so. Given its appeal to the notion of an epistemic entitlement, this two-part answer to the achievement problem is not happily regarded as depending on the sufficiency (for knowledge) of mere cross-world reliability. (Goldberg (2006), p. 304)

[85] Where does the entitlement derive from? And what makes it capable of surviving such environmental

La cita es clara. No entraremos a explicar (o intentar entender) qué rol juegan "los juicios relevantes en el razonamiento crítico", o cómo podría esto aportar "legitimación" a nuestros juicios sobre nuestros estados mentales. Sólo queríamos hacer notar que el mismo Burge afirma que la legitimación epistémica de estos juicios no viene de su naturaleza fiable sino, entre otras cosas, de la relación constitutiva que hay entre el pensamiento de primer orden y el pensamiento *cogito*.

Por supuesto, Brown seguramente respondería que, incluso si eso es así, incluso si la respuesta no tiene nada que ver con la fiabilidad de los pensamientos *cogito*, igualmente los pensamientos *cogito* no tienen nada que ver con las capacidades discriminatorias (el problema no era la presunta naturaleza fiabilista de la respuesta de Burge, sino que esa respuesta no decía nada sobre capacidades discriminatorias). De nuevo, creemos que Brown falla en sus críticas a Burge. Éste explícitamente niega que, al menos si nos ceñimos al auto-conocimiento, uno necesita de capacidades discriminatorias para adquirir conocimiento (esto es, (Discp) sería falso):

> El hecho de que no podamos usar signos fenoménicos o investigación empírica para distinguir nuestros pensamientos de otros pensamientos que estaríamos pensando si hubiéramos estado en un entorno diferente en ningún modo mina nuestra capacidad para conocer cuáles son nuestros pensamientos. "Individuamos" nuestros pensamientos, o los distinguimos de otros, pensando aquéllos y no otros, auto-adscriptivamente. (...) Para su justificación, el auto-conocimiento básico en ningún modo necesita ser complementado mediante investigaciones discursivas o comparaciones. (...) Pero hay diferencias fundamentales [entre el auto-conocimiento básico y la creencia perceptiva]. Un requerimiento de que, para conocer qué pensamientos estamos pensando, debamos primero ser capaces de distinguir nuestros pensamientos de pensamientos gemelos es, en mi opinión, incluso menos plausible que una posición análoga en cuanto al conocimiento perceptivo. [86]

Esto es, (Discp) es falso, y exigir capacidades discriminatorias para poder conocer que

---

switches? I think that the relevant entitlement derives not from the reliability of some causal-perceptual relation between cognition and its object. It has two other sources. One is the role of the relevant judgments in critical reasoning. The other is a constitutive relation between the judgments and their subject matter – or between the judgments about one's thoughts and the judgments' being true. (Burge (1996), p. 245)

[86] The fact that we cannot use phenomenological signs or empirical investigation to discriminate our thoughts from other thoughts that we might have been thinking if we had been in a different environment in no way undermines our ability to know what our thoughts are. We "individuate" our thoughts, or discriminate them from others, by thinking those and not the others, self-ascriptively. (...) For its justification, basic self-knowledge in no way needs supplementation from discursive investigations or comparisons. (...) But there are fundamental differences [between basic-self-knowledge and perceptual belief]. A requirement that, to know what thoughts we are thinking, we must be able first to discriminate our thoughts from twin thoughts is, in my view, even less plausible than the analogous position with regard to perceptual knowledge. (Burge (1988), p. 656)

p, cuando *p* es un juicio acerca de nuestros propios estados mentales, no está justificado. Si para Brown esto no supone una respuesta al argumento de la discriminación, nosotros no llegamos a ver qué sí lo podría suponer.

3.3. LA CRÍTICA DE BOGHOSSIAN

Uno de los pilares de la estrategia compatibilista de Burge (1988) es que el auto-conocimiento y el conocimiento empírico están regidos por principios epistémicos distintos (el suyo, pues, no entraría dentro de un modelo observacional del auto-conocimiento). Es más, si nos ceñimos a aquello que Burge denomina "auto-conocimiento básico", parece que al menos en esos casos podemos tener aquello que Boghossian llamaba "auto-conocimiento basado en nada": el mero hecho de tener un pensamiento *cogito* es suficiente para obtener auto-conocimiento, el auto-conocimiento básico no se basa en ningún tipo de evidencia. Por eso, Boghossian asume que Burge está proponiendo un modelo de auto-conocimiento que se enmarca dentro de este tercer modelo de "auto-conocimiento basado en nada".

Boghossian relacionaba este modelo de auto-conocimiento con aquello que él denomina "epistemología insustancial". En algunos casos es posible conocer proposiciones contingentes sin basarnos en evidencia alguna; el ejemplo más claro lo formaría nuestro conocimiento de proposiciones como "yo estoy aquí ahora". Este tipo de conocimiento se basa en una epistemología insustancial, y no supone ningún logro cognitivo. Al fin y al cabo, expuesto de manera un tanto vaga, parece que uno no necesita hacer demasiados méritos para llegar a conocer que está "aquí" y "ahora", y aunque tenga justificación de algún tipo para conocer tales cosas, esa justificación nada tiene que ver con inferencias que hace el sujeto desde su evidencia.

La cuestión es que, según Boghossian, no es aceptable que el auto-conocimiento se base en este tipo de epistemología insustancial, ya que no cumple las características típicas del conocimiento de este tipo. Identifica tres de estas características típicas: insensibilidad a la atención y a la dirección e infalibilidad—defiende que el auto-

conocimiento no cumple ninguna de estas condiciones, y que es la tercera característica, infalibilidad, la que incumple de forma más flagrante:

> La consideración más importante contra un modelo insustancial del auto-conocimiento (...): básicamente, que el auto-conocimiento es falible e incompleto. Tanto en el dominio de lo mental como en el de lo físico, pueden ocurrir eventos acerca de los cuales uno no sabe nada; y, en ambos dominios, incluso cuando uno se percata de la existencia de un evento, uno puede errar a la hora de caracterizar su carácter, creyendo que tiene una propiedad que de hecho no tiene. ¿Cómo se puede explicar esto? No conozco ninguna alternativa convincente al siguiente tipo de explicación: la diferencia entre acertar y fallar (sea por ignorancia o por error) es la diferencia entre estar en una posición epistémicamente favorable con evidencia relevante – y no estarlo. Por ponerlo de otro modo, el único modo de explicar los fallos del auto-conocimiento es entendiéndolo como siendo un logro cognitivo.[87]

El auto-conocimiento básico de Burge no se basa en evidencia alguna, lo único que necesitamos para saber qué estamos pensando es pensar un pensamiento *cogito*. Y hay otro problema según Boghossian, y es que esta propuesta sólo puede explicar una mínima parte de nuestro auto-conocimiento. Por ejemplo, los pensamientos *cogito* no podrían explicar el conocimiento que tenemos de nuestros estados mentales *permanentes*:

> [El auto-conocimiento básico y los pensamientos *cogito*] en ningún modo explican nuestro conocimiento de nuestros estados mentales *permanentes*. Los juicios concernientes a tales estados (...) no son auto-verificantes. (...) sí conocemos nuestras creencias y deseos en un modo directo y autoritativo, y no parece que la propuesta de Burge tenga los medios para explicar cómo.[88]

Las creencias serían el paradigma de un estado mental *permanente*[89]. Obviamente, el que tengamos un pensamiento *cogito* que tiene como objeto un estado mental

---

[87] The most important consideration, however, against an insubstantial construal of self-knowledge (...): namely, that self-knowledge is both fallible and incomplete. In both the domain of the mental and that of the physical, events may occur of which one remains ignorant; and, in both domains, even when one becomes aware of an event's existence, one may yet misconstrue its character, believing it to have a property it does not in fact possess. How is this to be explained? I know of no convincing alternative to the following style of explanation: the difference between getting it right and failing to do so (either through ignorance or through error) is the difference between being in an epistemically favorable position with relevant evidence – and not. To put this point another way, it is only if we understand self-knowledge to be a cognitive achievement that we have any prospect of explaining its admitted shortcomings (Boghossian (1989), p. 167)

[88] [Basic self-knowledge and *cogito* thoughts do not] at all explain our knowledge of our *standing* mental states. Judgments concerning such states (...) are not self-verifying. (...) we do know about our beliefs and desires in a direct and authoritative manner, and Burge's proposal seems not to have the resources to explain how. (Boghossian (1989), p. 169)

[89] No entraremos a proponer una descripción concisa de lo que son los estados mentales *permanentes* y los estados mentales *momentáneos*. Mencionemos simplemente que parece bastante razonable intentar reducir los estados *permanentes* a propiedades disposicionales del sujeto, y los estados *momentáneos* a su realización concreta en un momento concreto. Sirvan como ejemplo de estados *permanentes* las

permanente no garantiza que sepamos que estamos en ese estado mental, ya que los pensamientos *cogito* de este tipo no son auto-verificantes. Supongamos que S no cree que *p*. S podría tener un pensamiento *cogito* "Creo que *p*" que, de hecho, es falso. Dado que hay pensamientos *cogito* sobre estados mentales *permanentes* que son falsos, se sigue que este tipo de pensamientos *cogito* no pueden ser auto-verificantes. Pero es más, Boghossian defiende que los pensamientos *cogito* también tienen una aplicación limitada incluso si nos ceñimos a estados mentales *momentáneos*:

> Los juicios sobre mí mismo acerca de qué deseo o temo en ese mismo momento, por ejemplo, son manifiestamente no auto-verificantes, en cuanto no necesito desear o temer nada en particular para poder juzgar que lo hago. (…) El mejor escenario posible para las propuestas de Burge incluyen un juicio sobre sí mismo acerca de un mero pensar o aprehender de una proposición. (...) E incluso en estos casos, el juicio será auto-verificante sólo si el juicio *coincide absolutamente* en el tiempo con el momento en el que es pensado el pensamiento sobre el que se hace el juicio.[90]

Esto es, los pensamientos *cogito* son auto-verificantes sólo cuando el estado mental de primer orden acerca del cual son es un estado mental formado por una actitud proposicional de pensar y el pensamiento pensado coincide en el tiempo completamente con el pensamiento *cogito*. Por lo tanto, los pensamientos *cogito* sólo nos pueden explicar una parte mínima de nuestro auto-conocimiento y, por eso, no pueden ser la base de una buena estrategia compatibilista.

La crítica de Boghossian a Burge, pues, tiene dos ejes principales: los pensamientos *cogito* de Burge sólo pueden explicar una parte mínima del auto-conocimiento y, además, la falibilidad del auto-conocimiento es un problema para cualquier propuesta que defienda que el auto-conocimiento no se basa en ninguna evidencia (querríamos decir que el auto-conocimiento es un logro cognitivo). Pero aunque está en lo cierto en estas cuestiones concretas que plantea, también es verdad que esto no mina la efectividad de la estrategia de Burge en cuanto respuesta compatibilista.

---

creencias, el miedo de Coraline a las arañas, o su deseo de que sus padres aprendan a cocinar; y como ejemplo de estados mentales *momentáneos* los dolores, la sensación de frío o calor, o la sensación concreta de miedo que siente Coraline en t al ver una araña peluda.

[90]Self-regarding judgments about what I occurrently desire or fear, for example, are manifestly not self-verifying, in that I need not actually desire or fear any particular thing in order to judge that I do. (…) The best possible case for Burge's purposes will involve a self-regarding judgment about a mere thinking or entertaining of a proposition (...) And even here, the judgment will only prove self-verifying if the time at which the judgment is made is *absolutely coincident* with the time at which the thought being judged about is thought. (Boghossian (1989), pp. 169-170)

No era intención de Burge explicar la naturaleza del auto-conocimiento basándose en sus pensamientos *cogito*; es un error enmarcarle dentro del tercer modelo de auto-conocimiento descrito por Boghossian. Burge afirma explícitamente que hay segmentos de nuestro auto-conocimiento que no se basan en juicios de este tipo; por ejemplo, Burge (1996) propone que nuestro conocimiento de nuestros estados mentales fenoménicos no intencionales (o, al menos, "no proposicionales") depende directamente de nuestro sistema perceptivo[91]. Así, dado que el mismo Burge dice explícitamente que no cree que todo el auto-conocimiento responda al modelo caracterizado por los pensamientos *cogito*, parece que no ha lugar para criticarle que hay segmentos de nuestro auto-conocimiento que estos pensamientos no pueden explicar (que es lo que hace Boghossian). Los pensamientos *cogito*, pues, tienen una presencia limitada entre los juicios sobre nuestros estados mentales, pero eso no va en detrimento de las opiniones de Burge; por contra, constituye una parte esencial de sus ideas.

Burge propone los pensamientos *cogito*, primero, como respuesta a los argumentos incompatibilsitas; su importancia reside en que blindan nuestro auto-conocimiento básico (es esto algo que acepta Boghossian[92]), y en su eficacia para constituir una respuesta compatibilista. Los argumentos incompatibilistas no tienen como objetivo sugerir que el externista tendría problemas para explicar la naturaleza del auto-conocimiento; después de todo, parece que el auto-conocimiento es un fenómeno complejo y difícil de explicar independientemente de cuál sea la teoría semántica que aceptemos. Estos argumentos dicen que, si el externismo es verdadero, entonces no sabemos qué estamos pensando (o al menos Oscar no sabría qué está pensando), porque no sabemos cuál es el contenido de aquello que estamos pensando (o creyendo o deseando). Y es justo este conocimiento, esta justificación, lo que blindan los pensamientos *cogito*—éstos muestran que uno no podría fallar a la hora de identificar *el contenido* de sus pensamientos actuales.

[91] "Opino que el conocimiento de nuestros dolores y demás sensaciones – en contraste con el conocimiento de nuestros estados y eventos proposicionales – es empírico en el sentido de que depende para su justificación en experiencias o creencias sensoriales. Los juicios que constituyen tal conocimiento son simplemente creencias sensoriales." ("I do think that knowledge of our pains and other sensations – as contrasted with knowledge of our propositional states and events – is empirical in the sense that it depends for its entitlement on sensory experience or sensory beliefs. Judgments that constitute such knowledge just are sensory beliefs." (Burge (1996), p. 254, nota a pie de página 13))

[92] Siempre que no tengamos en cuenta el argumento de la memoria, tal y como veremos en la segunda parte del trabajo.

La segunda mitad de la crítica de Boghossian concernía a la falibilidad. Como hemos dicho, los pensamientos *cogito blindan* nuestro auto-conocimiento básico y, parece, esto supone un problema porque, como dice Boghossian, nuestros juicios acerca de nuestros estados mentales son falibles, el auto-conocimiento es un logro cognitivo. De nuevo, creemos que estas críticas no son acertadas. Opinamos también aquí que es necesario tener en mente cuál es la función que desempeñan los pensamientos *cogito* en la estrategia compatibilista de Burge: *blindar* el conocimiento que tenemos de *los contenidos* de nuestros pensamientos. Y no parece en principio descabellado asumir que *este* conocimiento se base en una "epistemología insustancial": al fin y al cabo los pensamientos *cogito* demuestran que lo único que tenemos que hacer para saber que estamos pensando que *p* (y no que *q*) es pensar esos pensamientos reflexivamente. No parece que el conocimiento que pretenden blindar los pensamientos *cogito* sea conocimiento falible, las críticas que vayan en esta dirección fallan de nuevo a la hora de identificar cuál es el objetivo (limitado) de postular este tipo de pensamientos[93].

Estas cuestiones sobre la falibilidad de los pensamientos *cogito* abren la puerta a un debate interesante que al menos nos gustaría mencionar aquí; nos referimos a la discusión acerca de hasta qué punto es posible el "error bruto" dentro de los márgenes del auto-conocimiento. Mencionemos, primero, que Burge (1996) afirma que este tipo de errores están muy limitados, y que un modelo observacional abriría la puerta a demasiados errores brutos:

[93] Esta idea, que el conocimiento de contenidos no es falible, parece sugerir una interpretación concreta sobre cómo es posible que los juicios sobre qué creemos o deseamos sean falibles. Supongamos, con Burge, que los pensamientos *cogito* sirven como paradigma para explicar qué tipo de justificación epistémica tenemos cuando juzgamos algo como "Creo que *p*" o "Deseo que *p*". Cuando hacemos juicios falsos de este tipo, o cuando los juicios verdaderos de este tipo que hacemos no constituyen conocimiento, eso no sucede porque no conozcamos cuál es el contenido de nuestra actitud proposicional, porque fallamos al identificar ese contenido, o porque no tenemos justificación para creer que el contenido de ese estado mental es el contenido de que *p*. Lo que "falla" es nuestro juicio acerca de qué tipo de relación guardamos hacia ese contenido (que identificamos correctamente). Dicho de forma un tanto basta, parece que lo que es falible es nuestro "conocimiento de actitudes", no nuestro "conocimiento de contenidos". También McLaughlin y Tye (1998) apuntan, en una nota a pie de página, que ni el hecho de que algunas veces no sepamos qué actitud guardamos hacia cierto contenido *p*, ni el que nos sea difícil explicar cómo es que adquirimos este conocimiento tienen nada que ver con el externismo semántico: "Es algo ampliamente aceptado que hay motivos para el escepticismo sobre si tenemos acceso privilegiado hacia qué tipo de actitud proposicional guardamos hacia *P*. Dado que los motivos son completamente independientes de la discusión de si el externismo semántico es verdadero, no trataremos el tema de si tenemos acceso privilegiado a si creemos que *P*, deseamos que *P*, o parecidos." ("It is fairly widely acknowledged that there are grounds for skepticism about whether we have privileged access to what kind of propositional attitude we have towards *P*. Since the grounds are entirely independent of the issue of whether externalism is true, we shall not be concerned here with whether we have privileged access to whether we believe that *P*, desire that *P*, ord the like." (McLaughlin y Tye (1998), p. 350, nota a pie de página 3).

> Hay límites severos en cuanto a los errores brutos en los juicios sobre las actitudes proposicionales presentes, ordinarias y accesibles de uno. Un error bruto es un error que no indica ningún fallo racional ni mal funcionamiento en los individuos equivocados. (...) Pero los errores sobre cuáles son los pensamientos y actitudes de uno normalmente parecen conllevar algún mal funcionamiento o deficiencia racional.[94]
>
> Una consecuencia de interpretar todo el auto-conocimiento según el modelo observacional simple es que los errores brutos – errores que no tienen consecuencias en la racionalidad o en el funcionamiento correcto de los juicios de revisión – son siempre posibles.[95]

Esto es, brevemente, parece que a la observación le es esencial la posibilidad del "error bruto"; los hechos observados son ontológicamente independientes de la observación y, por eso, la observación podría caer en un error bruto. Ahora, si el auto-conocimiento es similar a la observación, aún sin estar claro en qué consiste esa similitud, parece que el objeto observado (el pensamiento de primer orden) será independiente de la observación y que, por eso, también en el auto-conocimiento siempre habrá posibilidad de error bruto. Y esto es un problema según Burge, ya que parece obvio que la posibilidad del error bruto tiene limitaciones severas respecto al auto-conocimiento; por ejemplo, no hay posibilidad de error bruto dentro de aquello que Burge llama 'auto-conocimiento básico'. Así, no parece claro que la discusión acerca de la falibilidad del auto-conocimiento favorezca al modelo observacional en detrimento del "auto-conocimiento basado en nada", y no parece que Boghossian tenga a mano un argumento convincente contra Burge (menos aún si tenemos en cuenta que Burge enmarca los pensamientos *cogito* dentro de un segmento concreto del auto-conocimiento).

Por todo ello, opinamos que la respuesta de Burge supone una buena respuesta al argumento incompatibilista; demuestra cómo podemos saber cuál es el contenido de nuestros estados mentales sin que haya posibilidad de error. Otra cosa es que los pensamientos *cogito* se puedan proponer como paradigma para explicar la naturaleza epistémica de nuestros juicios acerca de nuestros estados mentales intencionales, pero eso por supuesto queda fuera del alcance de este trabajo. Lo que nos interesa es que la de Burge es una buena respuesta a los argumentos de Boghossian y Brown.

---

[94] ...there are severe limits on brute errors in judgments about one's present ordinary, accessible propositional attitudes. A brute error is an error that indicates no rational failure and no malfunction in the mistaken individuals. (...) But errors about what one's thoughts and attitudes are normally seem to involve some malfunction or rational deficiency. (Burge (1996), p. 251)

[95] A consequence of interpreting all self-knowledge on the simple observational model is that in any given case brute errors – errors that do not reflect on the rationality or sound functioning of the reviewing judgment – are possible. (Burge (1996), p. 255)

4. FALVEY Y OWENS

Kevin Falvey y Joseph Owens ofrecen su respuesta al argumento incompatibilista en "Externalism, Self-Knowledge, and Skepticism" (1994). Ésta se reduce a defender que los argumentos incompatibilistas de este tipo se basan en un principio falso sobre conocimiento, evidencia y exclusión de alternativas relevantes. Presentaremos primero brevemente su respuesta, así como algunas críticas que se les han hecho—al final del capítulo defenderemos que Falvey y Owens marcan la dirección correcta.

4.1. FIABILIDAD Y EVIDENCIA EXCLUYENTE

Comienzan por especificar dos principios que imponen condiciones al conocimiento:

(RA) Si
(i) *q* es una alternativa relevante a *p*, y
(ii) la creencia de S de que *p* está basada en evidencia que es compatible con que sea el caso que *q*, entonces

S no sabe que *p*.

(RA') Si
(i) *q* es una alternativa relevante a *p*, y

(ii) la justificación que tiene S para creer que *p* es tal que, si *q* fuera verdadero, entonces S seguiría creyendo que *p*, entonces

S no sabe que *p*.[96]

Podríamos resumir (RA) como la tesis de que alguien tiene justificación para creer que *p* sólo si tiene evidencia suficiente para excluir los escenarios relevantes en los que no se da que *p*, y (RA') como la tesis que alguien tiene justificación para creer que *p* sólo si su medio de formación de creencias es *fiable*, sólo si en todo escenario relevante en el que no se da que *p*, tal proceso de formación de creencias no le llevaría a creer que *p*.

De acuerdo con Falvey y Owens, el argumento de Boghossian se basa en la aceptación del principio (RA)[97]. El quid del argumento es que la evidencia que obtendría mediante introspección la víctima de una transición lenta no basta para excluir un escenario relevante en el que está pensando un pensamiento distinto; esto es así porque en el escenario alternativo, al tener las mismas propiedades internas que tiene en el escenario actual, obtendría la misma evidencia introspectiva que de hecho tiene. Falvey y Owens argumentan que (RA) es falso y que, si parece plausible, eso es porque en algunos casos uno obtiene fiabilidad sólo si tiene evidencia que excluye los escenarios alternativos.

A primera vista, (RA) parece un principio plausible. Si la evidencia E es compatible tanto con que se dé que *p* como con que no se dé que *p* siendo los dos escenarios relevantes, la elección de un de los dos escenarios sobre la única base de E parecería arbitraria. Parece plausible pensar que, por ejemplo, el conocimiento empírico se rige según tales condiciones sobre conocimiento y evidencia. Así, Falvey y Owens defienden que (RA) impone una condición necesaria para el conocimiento de hechos externos basado en evidencia perceptiva:

> (RA) seguro expone una condición necesaria plausible para que la creencia de un sujeto constituya conocimiento, cuando la creencia es una creencia que concierne al mundo externo, como en el caso de creencias perceptivas.[98]

---

[96] (RA) If (i) *q* is a relevant alternative to *p*, and (ii) S's belief that *p* is based on evidence that is compatible with its being the case that *q*, then S does not know that *p*.
(RA') If (i) *q* is a relevant alternative to *p*, and (ii) S's justification for his belief that *p* is such that, if *q* were true, then S would still believe that *p*, then S does not know that *p*. (Falvey y Owens (1994), p. 116)

[97] También nosotros hemos caracterizado así el argumento en el primer capítulo.

[98] (RA) surely states a plausible necessary condition for a subject's belief to constitute knowledge, where the belief is a belief concerning the external world, such as a perceptual belief. (Falvey y Owens (1994), p. 116)

(RA) por ejemplo explicaría por qué nos sentimos inclinados a rechazar que alguien pueda obtener conocimiento en escenarios como el propuesto por Goldman (1976). En ese ejemplo, Tom (quien pasea por el campo) se encuentra con lo que de hecho es un granero. Sobre la base de su percepción, forma la creencia verdadera de que tiene un granero enfrente. Pero dado que los alrededores contienen varios graneros falsos que, sobre la base de su evidencia perceptiva, Tom erróneamente tomaría por graneros auténticos (la evidencia perceptiva de Tom no es suficiente para excluir un escenario en el que está viendo un granero falso), el ejemplo supone que tenemos la fuerte intuición de que no sabe, sobre la base de su evidencia perceptiva, que tiene un granero enfrente—y (RA) explica esa intuición.

Aun así, Falvey y Owens defienden que, aunque (RA) pueda ser verdadero si nos ceñimos a conocimiento basado en evidencia perceptiva, el principio no resulta generalizable, y además (RA) resulta plausible porque se basa en (RA').

> ¿Por qué mina, el que Tom carezca de evidencia que excluya la posibilidad de que el objeto sea un facsímile, su capacidad de poseer conocimiento? Intuitivamente, la respuesta es que, dada la presencia de facsímiles en las inmediaciones de Tom, Tom podría fácilmente *engañarse* a pensar que un facsímile era un granero auténtico. Concretamente, dado que la creencia de Tom de que el objeto es un granero se basa en la apariencia visual del objeto, si el objeto *fuera* un facsímile, Tom seguiría creyendo que era un granero auténtico. Parece por lo tanto, que la plausibilidad de (RA) se basa en un principio más básico, [(RA')][99]

Si (RA) resulta plausible cuando tratamos de conocimiento basado en evidencia perceptiva, eso es así porque en tales casos uno tiene un método *fiable* sólo si tiene evidencia suficiente para excluir las alternativas relevantes (en tales casos (RA) implica (RA')). Resulta difícil encontrar ningún argumento a favor de este carácter "más básico" de (RA') en el artículo de Falvey y Owens, parece que según ellos tenemos la intuición de que una evidencia es suficiente cuando es *fiable*, cuando no nos llevaría a engaño en un escenario contrafáctico relevante.

[99] Why does Tom's failure to possess evidence ruling out the possibility that the object is a facsimile undermine his claim to knowledge? Intuitively, the answer is that given the presence of the facsimiles in Tom's vicinity, Tom could easily be *deceived* into thinking that a facsimile was a genuine barn. In particular, since Tom's belief that the object is a barn is based on the object's visual appearance, if the object *were* a facsimile, Tom would still believe that it was a genuine barn. It seems then, that the plausibility of (RA) is grounded in a more basic principle, namely [(RA')]. (Falvey y Owens (1994), p. 116)

Sea como fuere, resulta que en el caso del auto-conocimiento, a pesar de que el antecedente de (RA) se cumple (la evidencia introspectiva de Oscar no es suficiente para excluir que está en la Tierra Gemela pensando acerca de bi-agua) el antecedente de (RA') no se cumple (su justificación es tal que, si estuviera en la Tierra Gemela no creería que está pensando que el agua se hiela a cero grados, creería que está pensando que la bi-agua se hiela a cero grados). Esto es, en tal caso, la incapacidad de Oscar de excluir que está en la Tierra Gemela pensando acerca de bi-agua no basta para concluir que su método es tal que le podría llevar a tener una creencia falsa; el método que tiene Oscar para llegar a creer que está pensando que el agua se hiela a cero grados es *fiable*. Así, una vez aceptamos que es (RA') (y no (RA)) el principio "más básico" que impone una condición necesaria para el conocimiento en general, se sigue que el externismo semántico no resulta una amenaza para el auto-conocimiento autoritativo.

> Se sigue que el externista puede abrazar sin peligro alguno la idea de las alternativas relevantes, tal y como se presenta en (RA'), de forma general, sin exponerse a la acusación de que el externismo que defiende mina el conocimiento introspectivo de contenido.[100]

Así, el externismo semántico no amenaza el auto-conocimiento autoritativo, y

> una lección que podemos aprender de la discusión precedente es que no se debería entender la introspección según el modelo perceptivo.[101]

4.2. CONTRA (RA) (Y (DISCP))

Varios autores (especialmente Brown (2004)) critican que Falvey y Owens no dan suficientes argumentos contra (RA). (RA) y (RA') imponen condiciones necesarias (no suficientes) para el conocimiento y, por eso, apostar por la plausibilidad de (RA') poco o nada podría hacer en contra de (RA)—nosotros creemos que Falvey y Owens (1994) al menos marcan una senda adecuada. En el capítulo 5, donde presentaremos la propuesta de McLaughlin y Tye (1998), explicaremos que los argumentos

[100] It follows that the externalist can safely endorse the relevant alternatives idea, as embodied in (RA'), quite generally, without opening herself to the charge that the externalism she espouses undermines introspective knowledge of content. (Falvey y Owens (1994), p.118)

[101] ...one lesson that can be learned from the preceding discussion is that introspection should not be thought of on the model of perception. (Falvey y Owens, (1994), p. 118)

incompatibilistas de Brown y Boghossian se basan en una noción internista de la evidencia[102]; en esta sección defenderemos que, sobre la base de esta interpretación de la evidencia, los principios como (RA) y (Discp) resultan falsos—propondremos un contraejemplo, que nada tiene que ver con transiciones lentas o semánticas externistas.

El ejemplo se basa en una película, *Cube*. La película comienza con una escena en la que cierto individuo, Sam, despierta en un habitáculo con forma de cubo. Sam no sabe dónde está, ni cómo ha llegado hasta allí. Después, en cuanto avanza la película, descubrimos que el habitáculo es parte de un edificio más grande que está formado por un número indeterminado (aunque alto) de habitáculos de este tipo, todos con exactamente la misma forma y dimensión, el mismo color y sin ningún tipo de muebles o adornos. Usemos letras y números para referirnos a los habitáculos, y usemos el nombre '1A' para referirnos al habitáculo donde ha despertado Sam. Como hemos dicho, Sam no sabe qué ha pasado, quién lo ha llevado a ese lugar, dónde está. Se pregunta a sí mismo dónde se encontrará. Y se responde ejemplificando un pensamiento que expresaría profiriendo el siguiente enunciado:

> 'Bueno, al menos sé que estoy aquí ahora.'

¿Es verdad lo que ha pensado Sam? ¿Sabe que está "aquí" "ahora"? Creemos que la respuesta evidente es que sí, que Sam sabe que está "aquí" "ahora" (de hecho, no podría ser de otro modo). Habrá quien critique que es un conocimiento superfluo, que dado que Sam no es capaz de identificar aquello que conoce como "aquí" con ningún sitio que conoce (por descripción o porque ya ha estado allí) es un conocimiento que no le servirá absolutamente de nada—y quien critique esto tendrá razón. Pero no nos importa: lo que nos interesa aquí es que parece innegable que Sam sabe que está "aquí" "ahora".

¿Por qué nos interesa el que Sam sepa que está "aquí" "ahora"? Bueno, quienes lo hayan llevado allí podrían haberlo dejado en otro habitáculo del edificio, en 3B, o en 7J; un escenario en el que Sam no se encuentra en 1A ("aquí") es relevante. Pero hemos dicho que todos los habitáculos son "idénticos"; las experiencias que hubiera tenido

[102] Con "noción internista de la evidencia" nos referimos a la asunción de que la evidencia que tiene uno sobreviene en sus propiedades internas, tal que si dos sujetos se encuentran en el mismo estado interno, entonces tienen la misma evidencia.

Sam si hubiera despertado en otro habitáculo en vez de en 1A son exactamente las mismas que las que tiene al despertarse en 1A. Por eso, una vez asumimos una noción internista de la evidencia, se sigue que la evidencia que tiene Sam es compatible con que se encuentre en 3B, con que no se encuentre "aquí". Del mismo modo, Sam no puede discriminar entre 1A y 3B[103]. Así, sobre la base de (RA) o (Discp) se sigue que Sam no sabe que está "aquí". Pero queremos mantener que Sam sabe al menos eso; por eso, concluimos que (RA) y (Discp) son falsos (al menos como principios generales que imponen condiciones a todo tipo de conocimiento).

El defensor de (RA) y (Discp) tiene una respuesta a mano: tan evidente como que Sam sabe que está "aquí" es que no sabe que está en 1A. Lo que Sam dice verazmente al proferir 'Sé que estoy aquí ahora' no es la proposición singular en parte formada por 1A (a la cual de hecho refiere la proferencia de 'aquí' hecha por Sam), ni un pensamiento dependiente del objeto al que refiere (al estilo de Evans (1982)), sino la proposición general en parte constituida por elementos descriptivos como *el lugar donde hago esta proferencia* (sea cual sea ese lugar)[104]. Esa proposición es verdadera, Sam sabe que está en el lugar donde está haciendo esa proferencia; pero la cuestión es que Sam diría verazmente la *misma* proposición en el escenario relevante en el que se encuentra en 3B y profiere el enunciado 'Sé que estoy aquí ahora'. Esto es, dado que los escenarios relevantes no son escenarios en los que aquello que pretende saber Sam es falso, (RA) y (Discp) no predicen que Sam no sabe que está "aquí", en el lugar donde está haciendo la proferencia. Por otro lado, en los escenarios relevantes sería falso que Sam está en 1A, por eso estos escenarios sí pueden minar el conocimiento de Sam de que está "aquí", en 1A. Pero eso no es ningún problema según el defensor de (RA) y (Discp), porque eso es algo que de hecho Sam no sabe. Lo que queremos decir cuando afirmamos que es evidente que Sam sabe que está "aquí" es que sabe que está en el lugar donde está haciendo la proferencia (sea cual sea ese lugar).

[103] Sam no tiene a mano evidencia alguna que le permita activar el conocimiento de que el habitáculo 1A no es el habitáculo 3B; Sam no puede discriminar entre 1A (bajo el modo de presentación según el cual se le presenta: su percepción de 1A) y 3B (bajo el modo de presentación según el cual se le presentaría si estuviera allí: su contrafáctica percepción de 3B).

[104] Si adoptáramos una semántica bi-dimensionalista, afirmaríamos que la proferencia de Sam lleva asociadas dos proposiciones: la proposición más o menos general de que *sé que estoy en el lugar donde hago esta proferencia en el momento de hacer esta proferencia* (decimos "más o menos general" porque la proposición es singular: está en parte constituida por la proferencia) y la proposición singular de que *sé que estoy en 1A en t*. La primera proposición es verdadera, la segunda falsa.

Creemos que la respuesta esbozada en el párrafo anterior es errónea. Cuando presentamos el ejemplo de Sam y mantuvimos que sabe que está "aquí" queríamos decir que Sam sabe que está "aquí" *de re*[105]. Evidentemente Sam sabe que está en el lugar donde ha hecho la proferencia, pero creemos que también es verdad que Sam sabe que está "aquí", en el sitio donde se encuentra (*de re*), en ese mismo lugar al que puede referir mediante ostensión.

Carecemos de una argumentación que demuestre que Sam sabe que está "aquí" *de re*. Parece bastante evidente que podemos tener pensamientos singulares acerca de los objetos que percibimos, por mucho que nunca antes nos hayamos topado con ellos y por mucho que no seamos capaces de diferenciarlos de otros objetos (con los que nos podríamos haber topado) si se nos presentan bajo según qué modo de presentación, y que podemos obtener conocimiento *de re* acerca de esos objetos que percibimos. Sam percibe el habitáculo 1A. Creemos que sería *ad hoc* negar que puede tener pensamientos acerca de 1A (*de re*), un lugar que está percibiendo y al cual puede referirse mediante ostensión[106]. De hecho, Sam cumple con las condiciones que la mayoría de los filósofos imponen para poder tener pensamientos singulares[107]. Creemos que es totalmente *ad hoc* negar que Sam puede saber que está "aquí" *de re*; además, es falso.

Esto demuestra que el externista (singular) no debería aceptar al mismo tiempo una noción internista de la evidencia y principios como (RA) y (Discp) (entendidos como principios generales que imponen condiciones a toda instancia de conocimiento). Puede

[105] Mencionemos que de lo que acabamos de sostener no se sigue que una proferencia del enunciado 'Sam sabe que está en 1A' expresaría una proposición verdadera; depende de las ideas sobre semántica de adscripciones de actitudes proposicionales que adopte uno. Si aceptáramos una teoría de indéxico oculto (como la defendida por Crimmins y Perry (1989)), por ejemplo, se seguiría que la proferencia es falsa; si aceptáramos una teoría de "solución pragmática" (como las de Salmon (1986) o Soames (1988, 2002)), que es verdadera (aunque pragmáticamente inapropiada).

[106] En una nota a pie de página, Burge (1988) apunta en la misma dirección: "Sé que estoy aquí (compárese: en la Tierra) en vez de en algún otro sitio (compárese: en la Tierra Gemela). Mi conocimiento va más allá de conocer que estoy en cualquier lugar donde esté. Tengo la capacidad normal de percibir y pensar sobre mi entorno. Tengo este conocimiento porque percibo mi entorno y no ningún otro entorno concebible, y lo tengo incluso en el caso en el que sean concebibles otros sitios que podría no distinguir mediante percepción o descripción de aquí." ("I know that I am here (compare: on earth) rather than somewhere else (compare: twin earth). My knowledge amounts to more than knowing I am wherever I am. I have normal ability to perceive and think about my surroundings. I have this knowledge because I perceive my surroundings and not other conceivable surroundings, and I have it even though other places that I could not distinguish by perception or description from here are conceivable." (Burge (1988), p. 661, n. 9))

[107] Al menos cumple las condiciones que imponen autores como Donnellan (1977), Burge (1977), Evans 1982), Jeshion (2010), o Kaplan (1989).

que estos principios resulten plausibles en algunos contextos concretos, especialmente si nos ceñimos al conocimiento obtenido basándonos en evidencia perceptiva, pero desde luego no son generalizables.

4.3. SERRÍN Y CEREBRO: RESPUESTA A UNA OBJECIÓN

Para terminar este capítulo, nos gustaría presentar y responder una objeción propuesta por McLaughlin y Tye (1998) a las ideas de Falvey y Owens. Según ellos, dado que (RA) y (RA') son compatibles, el defensor de (RA) siempre podrá defender que, independientemente de que (RA') sea verdadero, también lo es (RA), y argumentan a favor de esta idea proponiendo ejemplos protagonizados por sujetos que, intuitivamente, no saben que *p*, tal que (RA) puede explicar por qué no lo saben mientras que (RA') no.

Presentan el caso de Oscar, quien cree (verazmente) que tiene cerebro. Una alternativa a esa creencia es que Oscar tenga la cabeza llena de serrín, y esa alternativa es relevante porque ciertos rumores en la aldea de Oscar afirman que un brujo ha cambiado los cerebros de algunos habitantes de la aldea por serrín. Oscar desconoce que si no tuviera cerebro no podría tener pensamientos o conducta y, por eso, se pregunta a sí mismo si él será una de las víctimas del brujo—para averiguarlo, consulta las entrañas de una gallina. Basándose en su interpretación de las entrañas, llega a la conclusión de que él no es una de las víctimas del brujo.

> Ahora, si la cabeza de Oscar estuviera llena de serrín, no seguiría creyendo que tiene cerebro; porque si su cabeza estuviera llena de serrín, no podría tener absolutamente ninguna creencia. Así, en este caso, el antecedente de (RA') no se satisface: dado que la alternativa relevante es falsa, su justificación no es tal que la alternativa relevante es verdadera. Por lo tanto, (RA') no dice nada acerca de si Oscar sabe que tiene cerebro. Pero claramente Oscar no lo sabe. Aquellos que proponen (RA) insistirán en que explica por qué no lo sabe. (...) Es más, [(RA')] no dirá nada sobre ningún caso en el que la creencia de S de que *p* depende contrafácticamente de *p*. Pero el defensor de (RA) insistirá en que no todos los casos de este tipo son casos de conocimiento de que *p*, y que (RA) explica por qué.[108]

[108] Now, if Oscar's head were full of sawdust, he would not still believe that he has a brain; for were his head full of sawdust, he would not have any beliefs at all. Thus, in this case, the antecedent of (RA') is not satisfied: since the relevant counterfactual is false his justification is not such that the relevant counterfactual is true. So, (RA') is silent about whether Oscar knows that he has a brain. But clearly

Del ejemplo se sigue, según McLaughlin y Tye, que, independientemente de nuestras opiniones acerca de (RA'), (RA) es un principio plausible, y que algo se perdería si lo rechazáramos.

Creemos que la objeción no es concluyente. Es verdad que (RA') no nos dice nada acerca de por qué Oscar no sabe que tiene cerebro, y que (RA) nos podría dar una explicación, pero no es la única posible. De hecho, Falvey y Owens podrían proponer una explicación del ejemplo más acorde con su defensa de (RA'). Ya hemos mencionado que el objetivo de (RA') es imponer una condición de fiabilidad al conocimiento: si la justificación que tiene S para creer que *p* no es fiable, entonces S no sabe que *p*. Algunos autores han distinguido entre lo que se denomina fiabilidad *local* y *global*:

> La fiabilidad global es fiabilidad para todos (o muchos de) los usos del proceso, no sólo para su uso en la formación de la creencia en cuestión. La fiabilidad local concierne sólo a la fiabilidad del proceso en el contexto de la creencia bajo consideración.[109]

Brevemente, un proceso mediante el cual se ha formado la creencia de S de que *p* será localmente fiable sólo si no hay ningún escenario relevante en el que *p* es falso y ese mismo proceso ha llevado a S a creer que *p*. Será globalmente fiable sólo si para un gran número de creencias, si esas creencias fueran falsas, entonces el proceso no hubiera llevado a los individuos en cuestión a esas creencias. (RA') impone una condición de fiabilidad *local*; es el hecho de que S tuviera la misma creencia de que *p* en una situación alternativa relevante lo que mina el conocimiento de S de que *p*. Pero si entendemos (RA') como imponiendo una condición necesaria (no suficiente) al conocimiento, (RA') es compatible con algún principio que imponga condiciones (necesarias y no suficientes) de fiabilidad *global* al conocimiento. Tal condición explicaría por qué Oscar en el ejemplo de McLaughlin y Tye no sabe que tiene cerebro. Las lecturas de entrañas de gallina no son *globalmente* fiables; llevan al sujeto que forma creencias basándose en esas lecturas a un gran número de creencias falsas. Así, a

---

Oscar does not know. Proponents of (RA) will insist that it explains why he does not know. (...) Indeed, [(RA')] will be silent about any case in which S's belief that *p* is counterfactually dependent on *p*. But a proponent of (RA) will insist that not all such cases are cases of knowledge that *p* and that (RA) explains why. (McLaughlin y Tye (1998), pp. 356-357, n. 15)

[109] Global reliability is reliability for all (or many) uses of the process, not just its use in forming the belief in question. Local reliability concerns only the reliability of the process in the context of the belief under assessment. (Goldman (1986), p. 45)

pesar de que (RA') no diga nada acerca del ejemplo de McLaughlin y Tye, es compatible con otras condiciones que sí diagnostican correctamente el ejemplo, y los cuales están en mayor consonancia con (RA') que (RA). Por eso, el ejemplo de McLaughlin y Tye no supone un gran argumento en favor de (RA); no vemos motivos suficientes para afirmar que (RA) nos es indispensable.

5. MCLAUGHLIN Y TYE

McLaughlin y Tye ofrecen en "Is Content Externalism compatible with Privileged Access?" (1998) la que opinamos es, junto con Falvey y Owens (1994), la respuesta más convincente al argumento. Proponen que el argumento de Boghossian se reduce a un "cuadrado incompatible", que el externismo semántico ni siquiera supone una amenaza *prima facie* para el acceso privilegiado, y que el externista puede mantener a la vez (RA) y la tesis de acceso privilegiado.

5.1. EL "CUADRADO INCOMPATIBLE"

Según McLaughlin y Tye (1998), el argumento incompatibilista de Boghossian viene a identificar un "cuadrado incompatible". Este "cuadrado" está formado por cuatro tesis que llevan a contradicción: la tesis de acceso privilegiado, una tesis defendiendo que el auto-conocimiento se basa en evidencia introspectiva, (RA), y la "Tesis de Pensamientos Alternativos":

> *Acceso Privilegiado*. Cuando nuestra facultad de introspección funciona adecuadamente, podemos conocer qué estamos pensando mediante introspección.[110]
>
> (RA) Si (i) *q* es una alternativa relevante a *p*, y (ii) la creencia de S de que *p* está basada en evidencia que es compatible con que sea el caso que *q*, entonces S no sabe que *p*.
>
> *Tesis de Evidencia Introspectiva*. El conocimiento introspectivo de qué estamos pensando en ese mismo momento se basa en evidencia que podemos adquirir mediante introspección.[111]
>
> *Tesis de Pensamientos Alternativos*. Para al menos un tipo de pensamientos momentáneos de que *P*, hay alguna circunstancia posible en la cual alguien está pensando que *P* en ese mismo momento, su facultad introspectiva está funcionando correctamente, y la evidencia que puede adquirir mediante introspección es compatible con que sea el caso que esté pensando algún pensamiento alternativo relevante de que *Q*.[112]

El cuadrado es incompatible: si la Tesis de Evidencia Introspectiva, la Tesis de Pensamientos Alternativos, y (RA) son verdaderos, se sigue que uno no puede saber mediante introspección qué está pensando (negación de Acceso Privilegiado). Es fácil ver cómo. Supongamos que la Tesis de Pensamientos Alternativos es verdadera, y que en algún caso S piensa que *p*, pero que la evidencia que tiene es compatible con que esté pensando que *q* (una alternativa, por lo que sea, relevante). Su creencia de que está pensando que *p* se basa sobre esa evidencia introspectiva que es compatible con el escenario alternativo relevante (Tesis de Evidencia Introspectiva). Sobre la base de (RA), se sigue que S no sabe qué está pensando, lo que supondría un contraejemplo a la tesis de Acceso Privilegiado.

Ahora, en principio el externismo semántico nada tiene que ver con este cuadrado. La cuestión es que, si se han propuesto escenarios de transición lenta, eso ha sido para argumentar que el externismo tiene como consecuencia la Tesis de Pensamientos Alternativos. Centrémonos en Oscar. Según el externista éste está pensando que el agua se hiela a cero grados, pero la evidencia a la que puede acceder mediante introspección es la misma que obtendría si estuviera pensando que la bi-agua se hiela a cero grados

[110] *Privileged Access.* When our faculty of introspection is functioning properly, we can know what we are thinking by introspection. (McLaughlin y Tye (1998), p. 350)
[111] *The introspective Evidence Thesis*. Introspective knowledge of what we are occurrently thinking is based on evidence that we can introspect. (McLaughlin y Tye (1998), p. 358)
[112] *The Alternative Thoughts Thesis*. For at least one type of occurrent thought that *P*, there is some possible circumstance in which one is occurrently thinking that *P*, one's faculty of introspection is functioning properly, and the evidence that one can introspect is compatible with its being the case that one is thinking some relevant alternative thought that Q. (McLaughlin y Tye (1998), p. 358)

(por contra, el internista nos dirá que Oscar está pensando lo mismo en los dos escenarios, ya que tiene las mismas propiedades internas en ambas ocasiones). Es así que el externismo semántico resulta una amenaza para el acceso privilegiado, a diferencia del internista, el externista está comprometido a asumir que hay ejemplos que resultan ser instancias de la Tesis de Pensamientos Alternativos y, por eso, está comprometido a rechazar alguna de las otras tres tesis del cuadrado incompatible (entre ellas la tesis de Acceso Privilegiado).

McLaughlin y Tye (1998) niegan que el externismo implique la Tesis de Pensamientos Alternativos, y dicen que de hecho tal tesis es falsa. De acuerdo con ellos, el externismo tiene tales consecuencias sólo si también se acepta cierta noción errónea de la evidencia introspectiva, es más, toda teoría aceptable del contenido tiene las mismas consecuencias una vez se acepta esa noción de la evidencia introspectiva. Por eso, concluyen, el externismo semántico no amenaza la Tesis de Acceso Privilegiado (o no al menos en mayor medida que cualquier teoría semántica aceptable).

Pero vayamos por partes. Primero, según McLaughlin y Tye, la siguiente tesis es una condición necesaria de la Tesis sobre Pensamientos Alternativos:

> *Tesis de la Infradeterminación*: Para al menos un tipo de pensamientos de que *P*, el que uno esté pensando que *P* (en *t*) no sobreviene en la evidencia que le es accesible mediante introspección (en *t*)[113]

La Tesis de Pensamientos Alternativos implica la Tesis de la Infradeterminación: si todo pensamiento de S sobreviniera en la evidencia introspectiva de S, difícilmente podría cierta evidencia introspectiva de S ser compatible con dos pensamientos distintos. Pero, según McLaughlin y Tye, la Tesis de la Infradeterminación sólo se sigue si aceptamos cierta teoría concreta (de hecho falsa) sobre la naturaleza de la evidencia introspectiva. De acuerdo con esta teoría, la evidencia que obtenemos mediante introspección se reduce a ciertos elementos puramente cualitativos. Así,

> La evidencia introspectiva que tiene uno en favor de que está pensando que *P* es que está teniendo una imagen auditiva con ciertas características cualitativas. ¿Qué son las características cualitativas en cuestión? Presumiblemente, incluyen características fonológicas imagísticas y de acento. (...) Si el hecho de que esté

[113] *The Underdetermination Thesis:* For at least one type of occurrent thought that *P*, whether one is occurrently thinking that *P* (at *t*) fails to supervene on the evidence introspectively available to one (at *t*). (McLaughlin y Tye (1998), p. 360)

> teniendo una imagen con ciertas características fonológicas imagísticas, de acento, y sintácticas *agota* la evidencia introspectiva disponible para Oscar después de sus transiciones cuando piensa que el agua [se hiela a cero grados], entonces, sin duda alguna, la evidencia introspectiva a su disposición falla al excluir la alternativa relevante de que está pensando que la bi-agua [se hiela a cero grados]. Por ello, si la evidencia en cuestión agota la evidencia introspectiva a disposición de Oscar, entonces no hay duda de que el externismo implica la Tesis de Pensamientos Alternativos.[114]

Esto es, supongamos que tal noción acerca de la evidencia introspectiva es verdadera, que la evidencia que obtenemos mediante introspección se reduce a ciertas propiedades puramente cualitativas de nuestros estados mentales. Si esto es así, evidentemente el externismo semántico implicará la Tesis de la Infradeterminación, ya que según el externismo nuestros pensamientos no sobrevienen en las propiedades cualitativas de nuestros estados mentales (estas propiedades cualitativas se individúan *internamente*). Del mismo modo, si aceptamos esta teoría sobre la evidencia introspectiva, el externismo semántico también implicará la Tesis de Pensamientos Alternativos; la víctima de una transición lenta sería un buen ejemplo: las propiedades cualitativas del estado mental de Oscar cuando piensa que el agua se hiela a cero grados son idénticas a las que instanciaría si estuviera pensando que la bi-agua se hiela a cero grados.

La cuestión es que, si aceptamos esta noción de evidencia introspectiva, entonces toda teoría semántica aceptable tendrá como consecuencias la Tesis de la Infradeterminación y la Tesis de Pensamientos Alternativos:

> De todos modos, el primer punto que nos gustaría señalar es que *si* la evidencia introspectiva de Oscar consiste en que tiene cierta imagen con estas características fonológicas imagísticas, de acento y sintácticas, entonces, para *cualquier* teoría sobre el contenido de los pensamientos que merezca consideración, la Tesis de la Infradeterminación será verdadera. Porque para cualquier teoría que merezca consideración, el pensamiento de Oscar de que el agua es líquida no sobrevendrá en la evidencia introspectiva a su disposición. La razón de ello es que cualquier teoría del contenido que merezca consideración será "externista" al menos en el siguiente sentido débil: implicará que el contenido de un estado de imagen auditiva no sobrevendrá en las características cualitativas en cuestión.[115]

[114] One's introspective evidence that one is thinking that *P* is that one is having an auditory image with certain qualitative features. What are the qualitative features in question? Presumably, they include imagistic phonological and stress features. (...) If the fact that he is having an image with certain imagistic phonological, stress, and syntactic features *exhausts* the introspective evidence available to travelling Oscar when he thinks that water is a liquid, then, to be sure, the introspective evidence available to him fails to rule out the relevant alternative that he is thinking that twater is a liquid. Thus, if the evidence in question exhausts the introspective evidence available to Oscar, then there is no question that externalism implies the alternative thoughts thesis. (McLaughlin y Tye (1998), p. 361)

[115] The first point we wish to stress, however, is that *if* Oscar's introspective evidence consists of his having an image with these imagistic phonological, stress, and syntactic features, then, on *any* theory of

> Dada esta noción de evidencia introspectiva, por lo tanto, cualquier teoría razonable sobre contenido implicará la Tesis de Pensamientos Alternativos.[116]

Una vez aceptamos que la evidencia introspectiva sobreviene en ciertas propiedades cualitativas de nuestros estados mentales, la única teoría semántica que no implicara la Tesis de Pensamientos Alternativos sería aquélla que defendiera que las propiedades de contenido de nuestros estados mentales sobrevienen en sus propiedades cualitativas. Pero una teoría del contenido tal no podría ser ni siquiera razonable para McLaughlin y Tye. Por eso, según ellos, si aceptamos tal noción de evidencia introspectiva, toda teoría mínimamente razonable acerca del contenido supondrá cierta amenaza para el defensor de la Tesis de Acceso Privilegiado, ya que éste verá reducidas sus opciones a negar o bien la Tesis de Evidencia Introspectiva o bien la tesis (RA). Y es exactamente por eso que McLaughlin y Tye niegan que el externismo semántico por sí mismo suponga amenaza alguna para la Tesis del Acceso Privilegiado, al menos no supone una amenaza mayor que cualquier teoría aceptable acerca de la naturaleza del contenido[117].

Así, McLaughlin y Tye defienden que debemos rechazar la noción de evidencia introspectiva que hemos presentado más arriba. Según ellos, no conocemos nuestros estados mentales inferencialmente, sino directamente; esto es, no es que "percibamos" cierto aspecto de nuestros estados mentales para luego inferir que estamos en tal estado mental, sino que tenemos acceso directo al estado mental en sí. Cuando Oscar se pregunta a sí mismo qué está pensando, no sucede que "perciba" ciertas propiedades cualitativas de su pensamiento de que el agua se hiela a cero grados y que luego infiera de esa evidencia puramente cualitativa que está pensando que el agua se hiela a cero grados; al contrario, Oscar accede mediante introspección a su pensamiento de que el agua se hiela a cero grados y basándose en esa evidencia, en ese pensamiento con el contenido de que el agua se hiela a cero grados, puede conocer que está pensando que el

---

thought content worthy of consideration, the underdetermination thesis will be true. For on any theory of content worthy of consideration, Oscar's thinking that water is a liquid will fail to supervene on the introspective evidence available to him. The reason is that any theory of content worthy of consideration will be "externalist" in at least this very weak sense: it will imply that the content of an auditory image state will not supervene on the qualitative features in question. (McLaughlin y Tye (1998), p. 361)

[116] Given this view of introspective evidence, then, any reasonable theory of content will imply the alternative thoughts thesis. (McLaughlin y Tye (1998), p. 362)

[117] Recordemos que Boghossian (1989) defendía que su argumento iba contra toda teoría relacionista del contenido, no sólo contra el externismo. Parece que McLaughlin y Tye opinan que toda teoría "mínimamente aceptable" sobre el contenido ha de ser relacionista; no creemos que sea tan relevante el énfasis en que las teorías semánticas aceptables pero no externistas amenazarían del mismo modo nuestro auto-conocimiento autoritativo.

agua se hiela a cero grados. El mismo estado mental constituye la evidencia sobre la que nos basamos para saber en qué estado mental nos encontramos.

> Intuitivamente, uno tiene acceso introspectivo a ciertos estados mentales, y estos estados mentales constituyen la evidencia de uno para sus creencias introspectivas, sin proveer una justificación proposicional para esas creencias. (...) los estados mentales que son la evidencia introspectiva del sujeto en tales casos son los pensamientos sobre los cuales son las creencias introspectivas. (...) Un defensor [del acceso privilegiado] puede mantener que Oscar tiene acceso introspectivo directo al mismo pensamiento que tiene en ese momento.[118]

Esta concepción de la evidencia introspectiva nos lleva a una importante disanalogía entre el auto-conocimiento y la percepción. Nuestro conocimiento empírico de hechos externos está mediado por las experiencias perceptivas que causan en nosotros esos hechos, *inferimos* nuestras creencias empíricas de nuestra evidencia perceptiva. Pero, a diferencia de lo que pasa en el conocimiento empírico, la relación entre nuestras creencias acerca de nuestros estados mentales y la evidencia sobre la que se basan no es *inferencial*, la relación es más bien *causal*. La relación evidencia-creencia es diferente en el caso del conocimiento empírico y en el del auto-conocimiento; el modelo observacional del auto-conocimiento es falso.

Dado que el mismo pensamiento de que *p* forma la evidencia introspectiva de S para conocer que está pensando que *p*, la evidencia introspectiva de S no es compatible con que esté pensando que *q* (y no que *p*). El pensamiento de Oscar de que el agua se hiela a cero grados forma su evidencia introspectiva para conocer que está pensando que el agua se hiela a cero grados; no es posible, con está noción de evidencia introspectiva en la mano, que uno esté pensando que *p* y que su evidencia introspectiva sea compatible con que no esté pensando que *p*, ya que el mismo pensamiento de que *p* es parte de su evidencia—se sigue que la Tesis de Pensamientos Alternativos es falsa. Por eso, no parece que la tesis (RA) sea problemática para alguien con esta noción de la evidencia introspectiva: dado que la evidencia que obtenemos mediante introspección cuando estamos pensando que *p* no es compatible con que no estemos pensando que *p*, el antecedente de (RA) no se cumple en los casos de auto-conocimiento.

---

[118] Intuitively, one has introspective access to certain mental states, and these states constitute one's evidence for one's introspective beliefs without providing a propositional justification for those beliefs. (...) the mental states that are a thinker's introspective evidence in such cases are the occurrent thoughts the introspective beliefs are about. (...) A proponent can maintain that Oscar has direct introspective access to the occurrent thought itself. (McLaughlin y Tye (1998), p. 364)

Así, McLaughlin y Tye concluyen que uno puede mantener al mismo tiempo una tesis externista sobre el contenido, que el auto-conocimiento se basa en evidencia, (RA), y que tenemos acceso privilegiado a nuestros estados mentales; ello depende de que tengamos una noción adecuada de la evidencia introspectiva. De ese modo, el externismo semántico no supone ninguna amenaza al auto-conocimiento autoritativo, ya que la aceptación o no del externismo semántico poco (o nada) tiene que ver con estas cuestiones acerca de la naturaleza de la evidencia introspectiva. El argumento propuesto por Boghossian no demuestra que el externismo semántico y el auto-conocimiento autoritativo son incompatibles, porque se basa en una teoría falsa sobre evidencia introspectiva. Una vez que rechazamos esa teoría, el externismo semántico no supone ninguna amenaza para el auto-conocimiento autoritativo.

5.2. UNA DISYUNCIÓN INCÓMODA

Lo que más nos interesa de la reconstrucción del argumento de Boghossian que hacen McLaughlin y Tye es el énfasis que ponen en el peso que tiene en el argumento una noción extremadamente internista de la evidencia introspectiva[119]. Rechazan que una noción tal sea aceptable, y demuestran que, una vez aceptamos una noción mínimamente externista de la evidencia introspectiva (tal que ésta no sobrevenga en las propiedades fenoménicas del sujeto), el argumento no amenaza la compatibilidad entre el externismo semántico y el auto-conocimiento autoritativo.

Hemos dicho en el capítulo anterior que, si aceptáramos esta interpretación internista de la evidencia, los principios del tipo de (RA) y (Discp) no podrían ser verdaderos (al menos entendidos como principios que imponen condiciones al conocimiento en general). Recordemos que, en el ejemplo de Sam, éste sabía que estaba "aquí" y "ahora" a pesar de que su evidencia (individuada internamente) era compatible con un escenario relevante en el que no estaba "aquí". Pero individuemos la evidencia de Sam externamente. Si el conocimiento de que está "aquí" se basa en evidencia, ésta estará

[119] McLaughlin y Tye hablan de las "características fonológicas imagísticas, de acento y sintácticas" de los estados mentales. No entendemos bien qué son las "características fonológicas imagísticas" de un estado mental, menos aún sus "características de acento"; nosotros hablaremos de sus "propiedades puramente cualitativas", propiedades que sobrevienen en la fenomenología de un sujeto.

constituida por la evidencia perceptiva de Sam, al menos parte de la evidencia que tiene Sam para saber que está "aquí" será que puede ver que está "aquí". Y, si individuamos esa evidencia externamente, aquello que Sam percibe, el lugar donde se encuentra, será razonablemente parte constituyente de la evidencia de Sam. Y aún aceptando (RA) podemos decir que Sam sí sabe que está "aquí", ya que su evidencia (el lugar que percibe) no es compatible con el escenario alternativo relevante. Una vez individuamos la evidencia externamente, pues, principios como (RA) son *prima facie* plausibles.

Ahora, como defienden McLaughlin y Tye, si negamos que la evidencia introspectiva de Oscar sobreviene en las propiedades cualitativas de sus estados mentales, el argumento de Boghossian no podrá apuntar a ningún tipo de tensión entre el externismo y el auto-conocimiento autoritativo. Así, el defensor del argumento de Boghossian se encuentra ante una disyunción incómoda. O individuamos la evidencia externamente, o (RA) será falso (o, quizás, ambas a la vez); no puede ser que nos aferremos a una noción internista de la evidencia mientras afirmamos que principios como (RA) imponen condiciones a toda instancia de conocimiento. Pero el argumento de Boghossian necesitaba de estas dos asunciones (así lo demuestran McLaughlin y Tye, y así lo exponen las premisas 1. y 4. en la formalización del argumento que hemos propuesto en el primer capítulo). Por lo tanto, al menos una de las premisas o asunciones del argumento de Boghossian habrá de ser falsa.

Independientemente de que creamos que deberíamos rechazar (RA) o la noción internista de la evidencia, parece, pues, que el argumento de Boghossian difícilmente podrá demostrar que el externismo semántico y el auto-conocimiento autoritativo son incompatibles. Uno no puede aferrarse a la vez a una noción internista de la evidencia y a principios como (RA).

5.3. EL ARGUMENTO DE LA DISCRIMINACIÓN Y LA DISYUNCIÓN INCÓMODA

Lo que nos proponemos ahora es explicar en qué medida creemos que la respuesta de McLaughlin y Tye al argumento de Boghossian puede suponer una respuesta al

argumento de la discriminación. Y es que Brown (2004) protesta que nada dice sobre discriminación y que, por eso, en ninguna medida supone una buena respuesta a su versión del argumento.

Primero, las quejas de Brown no están del todo justificadas. Cuando en el segundo capítulo hemos presentado su argumento de la discriminación hemos dicho que una de las premisas en las que se basaba el argumento era que el externista estaba comprometido a negar que el contenido es transparente. En una nota a pie de página McLaughlin y Tye (1998) explícitamente niegan tal cosa:

> Falvey y Owens llaman la atención sobre un punto importante, sobre que la tesis del conocimiento introspectivo del contenido no implica *la tesis comparativa* de que "con respecto a cualesquiera dos de sus pensamientos o creencias, un individuo puede conocer autoritativa y directamente ... si tienen el mismo contenido o no". Esta tesis, mantienen, es incompatible con el externismo. Pero también mantienen que eso no es un problema para el externismo, ya que la tesis es falsa. No entraremos a discutir esta tesis comparativa en este artículo. Nuestra opinión, la cual esperamos elaborar en algún otro lugar, es que la tesis comparativa, *con ciertas cualificaciones*, es verdadera. También defendemos que, dadas estas cualificaciones, la tesis no presenta dificultad alguna para el externismo.[120]

No mencionan cuáles son esas "cualificaciones" y, hasta donde sabemos, no llegaron a "elaborar sus opiniones en otro lugar"[121], con lo que resulta difícil valorar esta nota a pie de página. Defender que el externismo y la transparencia del contenido son compatibles sí supone una respuesta al argumento de la discriminación; ahora, también es verdad que McLaughlin y Tye no desarrollan esta vía, y que no está claro en qué medida su respuesta a Boghossian puede tener que ver con su adhesión a la tesis de la transparencia.

Contra Brown, creemos que es posible presentar una disyunción análoga a la presentada en la sección anterior, así como una versión de la respuesta compatibilista de McLaughlin y Tye que iría en contra del argumento de la discriminación.

---

[120] Falvey and Owens make the important point that the introspective knowledge of content thesis does not imply *the comparative thesis* that "with respect to any two of his thoughts or beliefs an individual can know authoritatively and directly ... whether or not they have the same contents". The latter thesis, they maintain, is incompatible with externalism. They also maintain, however, that this is not a problem for externalism since the thesis is false. We shall not be concerned with the comparative thesis in the present paper. Our view, which we hope to elaborate elsewhere, is that the comparative thesis, *with certain qualifications*, is true. We also hold that, given these qualifications, the thesis presents no difficulty for the externalist. (McLaughlin y Tye (1998), p. 355, n. 14)

[121] Tye (1998), sí que da alguna pista sobre en qué medida cree que el externismo es compatible con la transparencia del contenido. Presentaremos y valoraremos este artículo en la tercera parte de este trabajo.

En la sección 2.2. de esta parte decíamos lo siguiente:

- Que uno discrimina entre *a* y *b* sólo bajo *modos de presentación*.
- Que discriminar entre *a* y *b* viene a ser "activar el conocimiento de que *a* y *b* son dos objetos distintos.
- Que hay una relación estrecha entre el modo bajo el cual se le presentan a uno dos objetos y la evidencia que tiene para creer que son distintos.
- Que el argumento de la discriminación presupone que el estado fenoménico de uno cuando piensa que *p* forma el modo de presentación del estado mental "pensar que *p*".

Bien, volvamos ahora al ejemplo de Sam presentado en el capítulo anterior. Hemos dicho que, si asumimos que el modo de presentación del habitáculo 1A sobreviene en el estado fenoménico de Sam, entonces éste no puede distinguir entre ese habitáculo y el habitáculo 3B y que, sobre la base de (Discp), concluimos que no sabe que está "aquí". Como queremos decir que Sam sí tiene este conocimiento, hemos sugerido que principios como (Discp) parecen falsos (al menos si pretenden imponer condiciones a todo tipo de conocimiento).

Pero uno podría argumentar que Sam sí puede distinguir entre 1A y 3B, que basta con que rechacemos la idea de que la fenomenología de Sam agota el modo en que se le presenta 1A. Varios autores (seguramente el más conocido sea Evans (1982)) han defendido que cuando uno percibe un objeto, puede distinguir entre ese objeto y cualquier otro. Así, según este criterio, dado que percibe 1A, Sam puede distinguir entre 1A y 3B y, aún aceptando principios como (Discp), podemos asumir que sabe que está "aquí". En cuanto los modos de presentación de objetos no sobrevienen en los estados fenoménicos de quien los percibe, los principios como (Discp) resultan *prima facie* aceptables.

Así, alguien podría intentar preservar la plausibilidad de principios como (Discp) para el ámbito del auto-conocimiento (como lo hacían McLaughlin y Tye con (RA)). Basta con que el modo de presentación de un pensamiento no sobrevenga en la fenomenología del sujeto. Un modo en principio plausible de hacer esto es proponiendo que hay cierta analogía entre la relación entre un sujeto y el objeto que percibe, y la relación entre un

sujeto y el pensamiento que piensa. Siguiendo con la analogía, si percibir un objeto basta para poder discriminar entre ese objeto y cualquier otro, pensar un pensamiento basta para poder discriminar entre ese pensamiento y cualquier otro (la analogía resulta más plausible si, con McLaughlin y Tye, aceptamos que tenemos cierto "acceso directo" a nuestros pensamientos). Pero, si esto es así, cuando Oscar piensa que el agua se hiela a cero grados puede discriminar entre ese pensamiento y cualquier otro, da igual cuánto haya viajado (y el contenido es, pues transparente); Oscar sabe mediante introspección qué está pensando, aun y cuando aceptamos principios como (Discp), si individuamos los modos de presentación de los pensamientos "externamente". En el segundo capítulo hemos dicho que el argumento de la discriminación presupone que S no podrá discriminar entre los escenarios W y W' mediante introspección si en esos escenarios S se encuentra en el mismo estado fenoménico; la respuesta "McLaughlin-Tye" al argumento de la discriminación de Brown diría que esa presuposición es falsa.

Y lo dicho en los párrafos anteriores sugiere otra "disyunción incómoda" que se le presenta al defensor del argumento de la discriminación. Uno no puede aferrarse a la vez a una "noción internista de los modos de presentación" y a (Discp): o individuamos lo modos de presentación "externamente" o (Discp) es falso. Al menos una de las presuposiciones o premisas del argumento de la discriminación habrá de ser, pues, falsa.

5.4. TRANSPARENCIA Y CAPACIDADES DISCRIMINATORIAS

Las dos disyunciones propuestas prueban que el argumento de Boghossian y el argumento de la discriminación poco podrán demostrar sobre la compatibilidad o no entre el externismo semántico y el auto-conocimiento autoritativo. Ahora, uno podría apostar por uno de los dos lados de la disyunción—así lo haremos en esta sección.

Creemos que hay motivos para guardar serias dudas acerca de que Oscar sí puede discriminar, mediante introspección, entre su escenario actual y el escenario alternativo relevante, o entre el pensamiento de que el agua se hiela a cero grados y el pensamiento de que la bi-agua se hiela a cero grados. Creemos por lo tanto que la fenomenología de

Oscar sí agota el modo de presentación de su pensamiento, que la evidencia introspectiva que tiene sobreviene en las propiedades cualitativas de sus estados mentales y, así, que principios epistémicos como (RA) y (Discp) son falsos como impuestos al auto-conocimiento—entre Falvey y Owens (1994) y McLaughlin y Tye (1998), nos quedamos con los primeros.

Hemos mantenido antes que creemos que hay una estrecha relación entre modos de presentación y evidencia, entre (RA) y (Discp). Hemos dicho, uno puede discriminar entre dos objetos (bajo los modos de presentación relevantes) sólo si la evidencia que tiene justifica la creencia de que esos dos objetos son distintos. El problema es que, si esto es así, no vemos cómo podría Oscar discriminar entre el pensamiento de que el agua se hiela a cero grados y el pensamiento de que la bi-agua se hiela a cero grados, y no basta con estipular que el acceso que tiene a aquello que está pensando es tal que Oscar puede discriminar entre ese pensamiento y cualquier pensamiento alternativo. Realmente no vemos cómo podría la evidencia introspectiva de Oscar justificar la creencia de que esos dos pensamientos son distintos, ya que si a Oscar se le presentaran esos dos pensamientos, no vendría a creer, mediante introspección, que son distintos. Creemos que el acceso a esos dos pensamientos no motivaría a Oscar a creer que son distintos y, por eso, creemos que no está en posición de discriminar entre ellos, y que el contenido no es transparente[122].

Además, como decía Brown (2004), Oscar no tiene las habilidades que se le deberían suponer si tuviera las capacidades discriminatorias que se le atribuyen. Por ejemplo, si en un momento ulterior le contáramos cómo ha sido su historia, sin decirle en qué entorno se encontraba en cada momento, sería incapaz de decirnos cuándo estaba pensando sobre agua y cuándo sobre bi-agua. Es extraño que a alguien se le presente *x* en t0, que pueda en t0 discriminar entre *x* e *y*, pero que en t1 no sea capaz de decirnos si aquello que se le presentó en t0 era *x* o *y* (cuando no ha olvidado nada y no ha recabado

[122] Alguien podría intentar salvar la transparencia del contenido y las capacidades discriminatorias de Oscar arguyendo que éste no puede "tener ante sí" los dos pensamientos relevantes, por ejemplo, porque las transiciones lentas son casos de reemplazo conceptual (creemos que es ésta la posición que defiende Tye (1998)). Será en la segunda y tercera parte del trabajo donde diremos algo acerca de si en las transiciones lentas hay cohabitación o reemplazo, no entraremos a discutir estas cuestiones ahora. Baste con decir que, como se verá, apostaremos a favor de la cohabitación conceptual y que, por eso, creemos que Oscar sí puede tener ante sí al mismo tiempo el pensamiento de que el agua se hiela a cero grados y el pensamiento de que la bi-agua se hiela a cero grados, y que, si se le preguntara, erróneamente opinaría que se trata del mismo pensamiento.

evidencia que mina su justificación anterior en t0). Por eso, apostamos por negar que Oscar puede discriminar entre pensamientos de agua y bi-agua mediante introspección, por negar que la evidencia introspectiva que tiene es incompatible con el escenario alternativo relevante.

Para terminar, queremos mencionar que el supuesto incompatibilista podría responder que, a pesar de que el ejemplo de Sam demuestra que (RA) es falso como principio general que impone condiciones a todo tipo de conocimiento, resulta plausible en algunos casos, y no se ha demostrado que no sea un principio plausible para instancias de auto-conocimiento. Es cierto. Por ejemplo, es nuestra opinión que (RA) es plausible como imponiendo condiciones al conocimiento empírico basado en evidencia perceptiva (individuada internamente).

Ahora, Boghossian no demuestra que es plausible que (RA) imponga condiciones al auto-conocimiento. Además, creemos, la cuestión es si el conocimiento de los contenidos de nuestros estados mentales se parece más a casos como el de Sam o a los casos corrientes de conocimiento empírico. Ya en el capítulo concerniente a Burge hemos defendido que los pensamientos *cogito* comparten algunas características con las instancias de aquello que Boghossian llamaba 'conocimiento basado en una epistemología insustancial', sobre todo, porque no son falibles. Lo dicho se puede extender a todo juicio acerca de cuál es el contenido de aquello que estamos pensando, y aquello que sabe Sam (que está "aquí" y "ahora") es el ejemplo paradigmático de conocimiento basado en epistemología insustancial. Así, opinamos razonable concluir que el conocimiento de los contenidos de nuestros estados mentales guarda ciertas semejanzas con el conocimiento basado en epistemología insustancial y que, por eso, (RA) no impone condiciones en este campo concreto.

6. LA PROPUESTA DE BROWN

La gran mayoría de las estrategias compatibilistas que hemos considerado aceptan que los casos de transición lenta proponen escenarios relevantes y que, por eso, en principio suponen una amenaza para la compatibilidad entre el externismo semántico y el auto-conocimiento autoritativo. Brown (2004), por contra, propone negar la relevancia de los escenarios alternativos que podrían minar nuestro auto-conocimiento. Estudia uno por uno el externismo social, de clases naturales y singular, y propone distintos argumentos para defender que los casos de transición lenta no pueden ser *normalmente* relevantes. Así, muestra cómo es posible que haya casos en los que el externismo semántico y el auto-conocimiento autoritativo son compatibles, y asegura que, en la gran mayoría de los casos, de hecho sabemos qué estamos pensando sobre la única base de nuestra evidencia introspectiva.

6.1. EXTERNISMO SOCIAL

De acuerdo con el externismo social, ciertas convenciones sociales (lingüísticas) adoptadas en nuestro entorno determinan en parte el contenido de nuestros pensamientos. Un modo de sugerir que el externismo social no es compatible con el

auto-conocimiento autoritativo es proponiendo transiciones lentas entre comunidades lingüísticas que emplean dos términos homofónicos pero con significados distintos. Brown argumenta que es imposible que tales transiciones sean masivas—enumera tres condiciones necesarias para que un caso de transición lenta de este tipo suceda:

> Centrémonos en esas características de la transición lenta que hacen que sea plausible que la transición lleve a un cambio en el pensamiento del cual [S] no se puede dar cuenta sin hacer uso de información empírica. Defenderé que tales características son: (1) [S] es víctima de una transición lenta entre dos comunidades lingüísticas que tienen en común una palabra pero la definen en un modo ligeramente diferente; (2) [S] ignora esta diferencia lingüística; y (3) aun así, es una hablante competente de los dos lenguajes y una usuaria competente de la palabra en cuestión en los dos lenguajes.[123]

Primero, señalemos que Brown acepta que en el mundo actual sí se pueden dar transiciones lentas entre distintas comunidades lingüísticas; Ludlow (1995a) ofrece varios ejemplos de términos del inglés que varían de significado dependiendo de que se usen en Gran Bretaña o los Estados Unidos o, como en el caso de 'pragmatist', en pequeñas comunidades lingüísticas dentro de la misma región; parece al menos posible que se den casos de transición lenta en el mundo actual[124]. Lo que niega Brown es que las transiciones sean masivas, porque las tres condiciones mencionadas no se pueden dar a la vez en un número alto de casos.

El argumento es simple: si los casos de transición lenta se hicieran masivos, entonces las condiciones (1) y (2) no podrían darse a la vez, ya que el único modo de que se preserve alguna diferencia entre los significados del término en cuestión (condición (1)) es que aquéllos que cambian de una comunidad a otra sean conscientes de esa diferencia (negación de la condición (2)). Parece evidente que, si las dos comunidades intercambiaran una cantidad considerable de miembros frecuentemente, la diferencia en los usos del término podría subsistir sólo si aquéllos que transitan supieran que hay una diferencia lingüística. Por eso, concluye Brown, los casos de transición lenta sólo pueden ser un fenómeno poco más que marginal.

---

[123] Let us focus on those features of the slow switch case that make it plausible that the switch leads to a change in thought that [S] cannot notice without making use of empirical information. I will argue that these features are: (1) [S] is slowly switched between two linguistic communities that share a single word but define it slightly differently; (2) [S] is ignorant of this linguistic difference; and (3) nonetheless, she is a competent speaker of both languages and a competent user of the target word in both languages. (Brown (2004), p. 139)

[124] Por supuesto, también podemos encontrar casos de este tipo en castellano: 'tenderete', 'carta', 'changurro' o 'realista' son algunos ejemplos.

6.2. EXTERNISMO DE CLASES

El ejemplo de Oscar que hemos venido discutiendo es un ejemplo de transición lenta que se puede proponer para argumentar que el externismo de clases naturales es incompatible con el auto-conocimiento autoritativo. Basándose en las condiciones necesarias para tener capacidades de reconocimiento relativas a clases naturales, Brown defiende que este tipo de transiciones no son posibles.

De acuerdo con Brown, uno puede tener un concepto de clase natural C sólo si es capaz de reconocer los ejemplares de c-s como c-s. Así, uno por ejemplo tendrá el concepto ORO sólo si es capaz de reconocer los ejemplares de oro, si es capaz de distinguir entre los objetos que son de oro de los que no lo son. Y, según Brown, estas capacidades de reconocimiento producen conocimiento; si alguien puede reconocer cierto ejemplar de C como un ejemplar de C, sabe que aquello que ha encontrado es un ejemplar de la clase C con la que se ha topado más veces. Pero,

> Supongamos que S tiene cierta capacidad de reconocimiento de una clase y, al encontrarse con un ulterior ejemplar de esa clase, piensa que es un ejemplar de la clase con la que se encontró antes. Esta creencia no constituiría conocimiento si hubiera una alternativa relevante en la que todo parece exactamente lo mismo pero en la que se encuentra un ejemplar de cierta clase-duplicado (...) Por eso, si sus capacidades de reconocimiento producirán conocimiento, no puede ser el caso que haya una situación alternativa relevante en la cual todo es lo mismo pero se encuentra un duplicado. En consecuencia (...) un sujeto tiene capacidades de reconocimiento para un objeto o clase *x* sólo si no hay un duplicado relevante de *x*.[125]

Y, evidentemente, esto tiene consecuencias directas en los casos de transición. Si las capacidades de reconocimiento que tenemos nos previenen de que haya duplicados relevantes de las clases para las cuales tenemos esas capacidades, será imposible formar

[125] Suppose that S has a recognitional capacity for a kind and, on encountering a further instance of the kind, thinks that it is an instance of the kind she previously encountered. This belief would not constitute knowledge if there were a relevant alternative situation in which everything seems the same but in which she encounters an instance of a duplicate kind (...) Thus, if her recognitional capacity is to be knowledge-yielding in the way described, it cannot be the case that there is a relevant alternative situation in which everything is the same but she encounters a duplicate. In consequence (...) a subject has a recognitional capacity for an object or kind *x* only if there is no relevant duplicate for *x*. (Brown (2004), pp. 143-144)

un escenario alternativo que sea relevante donde haya un duplicado de una clase que contiene nuestro entorno (tal que seamos incapaces de distinguir entre ellas).

> Este resultado sobre las capacidades de reconocimiento para clases tiene consecuencias para los pensamientos sobre clases basados en reconocimiento. Supongamos que un sujeto piensa el pensamiento basado en reconocimiento que eso (x) es F. Es parte de su capacidad de reconocimiento para x el que no haya duplicados relevantes. Así, no hay ninguna situación alternativa relevante en la que se encuentra un duplicado en vez de x. A fortiori, no hay ninguna situación alternativa relevante en la que se encuentra un duplicado, desarrolla una capacidad de reconocimiento para ese duplicado, y así piensa un pensamiento basado en reconocimiento sobre el duplicado.[126]

Brevemente: uno tiene un concepto de clase natural C sólo si tiene las capacidades de reconocimiento necesarias, y uno tiene tales capacidades de reconocimiento sólo si no hay ninguna clase natural relevante que confundiera con C. Por eso, si uno tiene el concepto de clase natural C, es imposible que un escenario que contenga una clase natural distinta pero indistinguible de C sea relevante. Así, concluimos que si Oscar realmente tiene el concepto AGUA, y está pensando que el agua se hiela a cero grados, entonces es imposible que un escenario en el que esté pensando que la bi-agua se hiela a cero grados (tal que Oscar no puede distinguir entre el escenario actual y el escenario alternativo) sea relevante. Dado que es imposible que tales escenarios sean relevantes, también es imposible que minen el auto-conocimiento autoritativo de nadie.

6.3. EXTERNISMO SINGULAR

La posición de Brown sobre el externismo singular es menos contundente que las que mantenía sobre el externismo social y el de clases naturales. Distingue entre los pensamientos singulares basados en reconocimiento y los pensamientos perceptivos basados en ostensión, y defiende que, aunque no se pueden plantear escenarios alternativos relevantes para los primeros, los ejemplos que plantean escenarios alternativos para los pensamientos basados en ostensión sí son más problemáticos.

[126] This result about recognitional capacities for kinds has consequences for recognition-based thoughts about kinds. Suppose that a subject thinks the recognition-based thought that that (x) is F. It is part of her having a recognitional capacity for x that there is no relevant duplicate. Thus, there is no relevant alternative situation in which she encounters a duplicate instead of x. A fortiori, there is no relevant alternative situation in which she encounters a duplicate, develops a recognitional capacity for that duplicate, and so thinks a recognition-based thought about the duplicate. (Brown (2004), p. 144)

Es fácil imaginarse una transición lenta que, en principio, sugiera que el externismo singular es incompatible con el auto-conocimiento autoritativo. Supongamos que *a* ve todos los días a las ocho de la mañana desde la ventana de su casa cómo *b* cruza la calle para ir a coger el metro. *a* no conoce a *b*, y decide llamarla por el nombre 'Señorita X'. Un día, *b* deja de cruzar la calle a las ocho de la mañana para ir a coger el metro, es su hermana gemela *c* la que sigue con la rutina que llevaba *b*. *a* no se da cuenta del cambio, y confunde a *c* con *b*, creyendo que la chica que ve ahora todos los días desde su ventana es la misma que ella bautizó como 'Señorita X'. Pasa el tiempo, y es *c* la que ahora cruza la calle todos los días para coger el metro; se supone que *a* ha tenido suficiente tiempo para adquirir el concepto C. Un día, *a* tiene un pensamiento que expresaría profiriendo el enunciado 'hoy la señorita X viste de rojo'; parece plausible concluir que *a* piensa (y cree) que *c* viste de rojo. Pero, según el incompatibilista, es un escenario relevante que *a* esté pensando que *b* viste de rojo. Dado que *a* no puede distinguir entre el escenario actual y el escenario relevante, dado que la percepción que tiene *a* de *c* no basta para que active el conocimiento de que *b* no es *c*, el incompatibilista concluye que *a* no sabe que está pensando que *c* viste de rojo.

Sabemos que Brown ya tiene una respuesta preparada. Hemos visto cómo niega que sea posible proponer escenarios alternativos relevantes para los pensamientos constituidos por conceptos de clase natural, los cuales exigen capacidades de reconocimiento. Los conceptos que expresamos al proferir nombres propios requieren las mismas capacidades de reconocimiento de acuerdo con Brown; así, si S tiene un pensamiento singular basado en reconocimiento (era el caso de *a* en el ejemplo presentado en el párrafo anterior) resulta imposible que haya un escenario alternativo relevante en el que S está pensando un pensamiento distinto, tal que el escenario alternativo y el actual sean indistinguibles para S.

La cosa resulta un poco más difícil si nos centramos en los pensamientos basados en ostensión. Brown admite que hay escenarios alternativos relevantes que minan nuestro auto-conocimiento de este tipo de pensamientos (aunque niega que estos casos sean frecuentes). Supongamos que S, trabajador de una empresa encargada de empaquetar productos agrícolas, observa una hilera interminable de manzanas en una cinta mecánica. S ve la manzana *a*, y piensa que esa manzana (*a*) es roja. Como dice Brown, "parece una alternativa relevante que S hubiera estado mirando una manzana-duplicado,

por ejemplo, si las manzanas hubieran llegado en la cinta en una orden ligeramente diferente"[127]. S no puede distinguir entre las dos situaciones, por eso, sobre la base de (Discp), se sigue que no sabe que está pensando que esa manzana (*a*) es roja. Brown concluye que sí es verdad que en algunos casos en los que pensamos un pensamiento perceptivo basado en ostensión no sabemos qué estamos pensando, a pesar de que esas situaciones no son muy comunes (la mayoría de los objetos sobre los cuales pensamos (personas, ciudades, mascotas, objetos personales) no tienen duplicados y para los objetos que es más plausible que sí lo tengan, parece que las condiciones para que un escenario alternativo sea relevante no se cumplen normalmente).

6.4. COMENTARIOS A LA PROPUESTA DE BROWN

Brevemente, pues, Brown defiende que el externismo semántico y el auto-conocimiento autoritativo son "normalmente compatibles", porque los escenarios alternativos que minan nuestro auto-conocimiento pocas veces son relevantes. Esto se debe a que:

1. Es difícil que se den las condiciones necesarias para que se dé una transición lenta, imposible que las transiciones lentas sean "masivas".
2. En cuanto a los pensamientos en parte constituidos por conceptos que requieren capacidades de reconocimiento (los conceptos de clase natural y los conceptos que expresaríamos usando nombres propios serían los casos paradigmáticos) es imposible que haya un escenario alternativo relevante.
3. A pesar de que es posible proponer escenarios alternativos relevantes para los pensamientos perceptivos basados en ostensión, esos casos también son extraños y minoritarios.

No creemos que esta estrategia sea acertada—mencionaremos tres motivos por los que entendemos que la propuesta es problemática: acepta e implica consecuencias que ya son indeseables, los criterios para tener conceptos de clase natural que presupone no son aceptables, y entiende mal la dialéctica de la discusión.

[127] ...it seems a relevant alternative that S might have been looking at a duplicate apple, say, if the apples had happened to come down the belt in a slightly different order. (Brown (2004), p. 146)

6.4.1. La bala incompatibilista.

Nos gustaría recalcar que Brown (2004) en parte "muerde la bala incompatibilista"; admite que, si se dan casos de transición lenta, entonces la víctima no puede conocer mediante introspección qué está pensando. Oscar, por ejemplo, no podría conocer mediante introspección que está pensando que el agua se hiela a cero grados—Brown no niega este punto, pero niega que esto tenga mayor importancia en la discusión acerca de la posible compatibilidad entre el externismo semántico y el auto-conocimiento autoritativo, ya que esos casos son puras ficciones filosóficas. En todo caso, acepta que el externismo semánticos al menos podría afectar al auto-conocimiento autoritativo de algún individuo, y esto ya es morder (aunque sea en parte) la bala incompatibilista.

Por otro lado, creemos que la propuesta de Brown ya tiene conclusiones indeseables. Acepta que es posible que en el mundo actual se den casos de transición lenta, lo que objeta es que no puede ser que estos casos sean masivos. Aceptar que el externismo semántico tiene la consecuencia de que algunos (pocos) individuos no pueden conocer mediante introspección qué están pensando es aceptar mucho—asumir eso es asumir que el externismo tiene consecuencias realmente indeseables y que, por lo tanto, hay motivos por los cuales el internismo semántico puede ser preferible.

Creemos que estas consecuencias son más palpables en los casos en los que uno piensa un pensamiento perceptivo basado en ostensión. Supongamos que, en la playa, S piensa, haciendo ostensión al grano de arena *a*, que eso (*a*) es un grano de arena. Es un escenario relevante que S hubiera hecho ostensión al grano de arena *b* porque, supongamos, S ha escogido el grano *a* al azar, y, por eso, es un escenario relevante que S estuviera pensando que eso (*b*) es un grano de arena. La evidencia perceptiva que tiene S no es suficiente como para que pueda distinguir entre *a* y *b* y, por eso, de acuerdo con el argumento de la discriminación, no sabe qué está pensando. Como hemos visto, Brown muerde la bala y asume que en una situación así S no sabría qué está pensando.

Pero esta posición que adopta Brown no parece tan fácil de asumir. Primero porque, de nuevo, no es deseable que nuestras teorías semánticas tengan estas consecuencias. Si el externismo semántico tiene la consecuencia de que cualquiera que en una playa piense

un pensamiento del tipo ESO ES UN GRANO DE ARENA haciendo ostensión a un grano concreto no sabe qué está pensando, parece que tenemos motivos para preferir una semántica internista. Ejemplos como el descrito parecen casos paradigmáticos de auto-conocimiento autoritativo, no queremos renunciar a ellos.

Por otro lado, la posición que adopta Brown tiene una consecuencia que, aunque a algunos quizás no les parezca problemática, nosotros sí creemos que es, en el mejor de los casos, extraña. Es una consecuencia de esta tesis (la tesis de que en la situación descrita S no sabe qué está pensando) que Brown (ni nadie) no puede saber que la tesis es verdadera (es más: Brown puede saber que no puede saber que la hipótesis que plantea es verdadera); lo contrario nos llevaría a contradicción. Expliquémonos.

Brown "podría forzarse" a ponerse en la situación de S. Así, puede decidir ir a la playa y pensar, haciendo ostensión a un grano de arena *a* que escoge al azar, un pensamiento del tipo ESTOY PENSANDO QUE ESO (*a*) ES UN GRANO DE ARENA, PERO NO SÉ QUE ESTOY PENSANDO QUE ESO (*a*) ES UN GRANO DE ARENA. Primero, si la hipótesis de Brown es verdadera, entonces lo que acaba de pensar Brown es de hecho verdadero: Brown está pensando que eso (*a*) es un grano de arena, pero no sabe que está pensando que eso (*a*) es un grano de arena. Supongamos ahora que Brown sabe que su hipótesis es verdadera: entonces sabe que lo que acaba de pensar es verdadero. Pero esto no puede ser, porque si sabe que lo que ha pensado es verdadero, sabe que está pensando que eso (*a*) es un grano de arena pero que no sabe que está pensando que eso (*a*) es un grano de arena. Por distribución del predicado "saber" a los dos lados de la conjunción, Brown sabe que está pensando que eso (*a*) es un grano de arena. Pero hemos dicho que, de acuerdo con la hipótesis de Brown, no sabía que estaba pensando que eso (*a*) es un grano de arena—hemos llegado a una contradicción[128]. Por lo tanto, Brown no puede saber que la tesis que propone es verdadera (no puede saber que S no sabría qué está pensando).

Puede que la hipótesis de Brown (que alguien en la situación de S no sabría qué está pensando) sea verdadera, pero es una hipótesis que nadie puede saber que es verdadera. Esto nos parece, en el mejor de los casos, extraño.

[128] Pongámoslo algo más claro. Lo que ha pensado Brown tiene la forma de que $p \wedge \neg Kp$, donde *p:* "estoy pensando que eso (*a*) es un grano de arena", y es verdadero de acuerdo con su hipótesis. Si supiera que lo que ha pensado es verdadero, se seguiría que $K(p \wedge \neg Kp)$ y, por distribución de K, que Kp y que $K\neg Kp$. Por lo tanto, tenemos Kp y $\neg Kp$, lo que supone una contradicción.

En la sección 2.3.3. de esta parte hemos presentado, muy brevemente, una estrategia compatibilista que se parecía en cierta medida a la propuesta de Brown. Esta estrategia decía que, dado que en las transiciones lentas son ejemplos de reemplazo conceptual, los escenarios alternativos nunca son relevantes y que, por eso, no pueden minar el auto-conocimiento autoritativo de los sujetos de las transiciones. Pero el escenario de S descrito arriba no es un ejemplo de transición lenta y, por eso, de nada puede servir acudir al reemplazo para negar que el escenario alternativo no es relevante—esto es, la estrategia basada en reemplazo no dice nada sobre si nuestro auto-conocimiento de los pensamientos perceptivos basados en ostensión está amenazado o no, pero debería decir algo. Como hemos visto, morder la bala y decir que en una situación tal no podríamos saber qué estamos pensando no parece una posición cómoda.

6.4.2. Reconocimiento y discriminación.

Pasemos ahora a las opiniones de Brown sobre auto-conocimiento de pensamientos en parte constituidos por conceptos que requieren capacidades de reconocimiento. Brown prácticamente se limita a afirmar que, si alguien piensa un pensamiento en parte constituido por un concepto tal, entonces es imposible que un escenario alternativo sea relevante, pero no desarrolla suficientemente las consecuencias de defender que tener un concepto de clase natural (o un concepto que expresaríamos profiriendo enunciados que contienen nombres propios) requiere estas capacidades de reconocimiento. Por ejemplo, lo que se sigue del argumento de Brown no es que, dado que Oscar está pensando que el agua se hiela a cero grados, y dado que AGUA es un concepto de clase natural (que requiere de capacidades de reconocimiento), entonces un escenario alternativo no puede ser relevante para Oscar, sino que, dado que un escenario alternativo es relevante para Oscar, entonces éste no tiene las capacidades de reconocimiento que se requieren para poseer el concepto AGUA y, por lo tanto, no puede pensar que el agua se hiela a cero grados.

Lo mismo sucedería en el caso de *a* descrito más arriba, quien conoce a las hermanas gemelas *b* y *c* por el nombre "la señorita X": se sigue de las opiniones de Brown que *a* nunca podría adquirir los conceptos B y C. Pero hay ejemplos parecidos a éste en el mundo real: no es extraño que alguien confunda a una persona por su gemela. Es más o menos frecuente que *x* conozca a *y*, que *y* tenga un hermano gemelo *z* (tal que si *x* se

encontrara con *z*, creería que es *y*), y que *x* desconozca que *y* tiene un hermano gemelo. *x* no tiene las capacidades de reconocimiento que le reclama Brown, se sigue que, hasta que descubra que *y* tiene un hermano gemelo, *x* no podrá adquirir ningún concepto Y. Pero esto es absurdo. Es a todas luces absurdo defender que *x* no podrá tener pensamientos *de re* sobre *y* hasta que descubra que *y* tiene un hermano gemelo.

Por eso, creemos que no es deseable defender que, para tener un concepto de clase natural C, alguien tiene que tener capacidades de reconocimiento (en el modo en el que Brown describe estas capacidades de reconocimiento). Es más, este requisito es incompatible con muchas de las vertientes externistas. Como vimos, el externista de clases naturales típicamente mantendrá que uno puede tener un concepto de clase natural C sin tener los medios para distinguir entre c-s y no-c-s (recuérdese la incapacidad de Putnam para distinguir entre hayas y olmos), y desde luego es incompatible con un externismo social que mantenga que es posible adquirir un concepto *deferencialmente*.

6.4.3. Dialéctica de la discusión.

Para terminar, creemos también que la respuesta de Brown falla a la hora de entender la dialéctica del argumento (al menos tal y como lo planteaba Boghossian); las respuestas compatibilistas basadas en negar la relevancia de las transiciones lentas no pueden ser una buena respuesta al argumento incompatibilista. Warfield (1992, 1997) ya defendía una posición parecida, y creemos que las respuestas de Ludlow (1995a, 1997) son concluyentes y decisivas (también contra las ideas de Brown).

Warfield (1992) defiende que el argumento de Boghossian (1989) tal y como está planteado es inválido. La conclusión del argumento es que el externismo semántico y el auto-conocimiento autoritativo son incompatibles y, según Warfield, lo máximo que se sigue de las premisas del argumento es que, si un caso de transición lenta como el descrito por Boghossian es actual, entonces la víctima de la transición no conoce mediante introspección qué está pensando.

> Pero esta conclusión no es relevante en cuanto a la cuestión de la compatibilidad entre el externismo y el auto-conocimiento introspectivo. Es relevante, *a lo sumo*, para la siguiente pregunta: Dado el externismo, es *necesario* que los contenidos de

> los pensamientos de un individuo sean cognoscibles para ese individuo sobre la base de la introspección?[129]

Ludlow (1995a) responde a Warfield (1992) que hay casos de transición lenta en el mundo actual, y que, de hecho, son bastante comunes (Brown está de acuerdo en que hay casos de transición lenta en el mundo actual aunque, seguro, no piensa que sean tan comunes como cree Ludlow). Además, defiende Ludlow (1995a), del hecho de que los casos de transición lenta sean comunes se sigue que, para un gran número de pensamientos que tenemos, un escenario alternativo (donde estamos pensando algo distinto y que es indistinguible para nosotros) se convierte en relevante; no es necesario el que de hecho seamos víctimas de una transición lenta para que un escenario alternativo nos sea relevante, basta con que las transiciones lentas sean frecuentes.

Nos gustaría abrir un pequeño paréntesis aquí. Creemos que Ludlow tiene razón en este punto; aun así, dejaremos de lado la discusión sobre si la frecuencia de las transiciones lentas es lo suficientemente grande como para que se siga que un escenario alternativo es generalmente relevante. Lo que queremos proponer con este paréntesis es que el grupo de individuos para los cuales un escenario alternativo es relevante es un poquito mayor de lo que creía Brown. El que un escenario alternativo sea relevante no depende sólo de factores externos, también puede depender de factores internos al sujeto (sus creencias, por ejemplo). Brown cree que en el mundo actual hay víctimas de transiciones lentas y, por lo tanto, sabe que hay probabilidades de que ella sea la víctima de una transición lenta. Para (prácticamente) cualquier pensamiento que piense Brown, si se pregunta a sí misma si habrá un escenario alternativo relevante que mina su auto-conocimiento de ese pensamiento, se sigue que tal escenario ya es relevante para ella. Creemos que de las ideas defendidas por Brown se sigue que, a nada que se plantee cuestiones de este tipo a sí misma (escenario, suponemos, nada extraño para ella), no podrá conocer mediante introspección qué está pensando. Cerremos el paréntesis.

Warfield (1997) responde a Ludlow (1995a), reiterando lo que ya defiende en Warfield (1992). Según Warfield (1997), Boghossian (1989) demuestra que es posible que el externismo semántico tenga como consecuencia que alguien no pueda conocer mediante

[129] But this conclusion is not relevant to the question of the compatibility of externalism and introspective self-knowledge. It is relevant *at most* to the following question: Q Given externalism is it *necessary* that the contents of a thinker's thoughts are knowable to the thinker on the basis of introspection? (Warfield (1992), p. 218)

introspección qué está pensando, y Ludlow (1995a) que hay un mundo posible concreto, el actual, en el cual el externismo semántico tiene esa consecuencia, pero ninguno de los dos demuestra que el externismo semántico y el auto-conocimiento autoritativo son incompatibles, que no hay ningún mundo posible en el cual las dos tesis son verdaderas.

Es la respuesta a este artículo dada por Ludlow (1997) la que más nos interesa. Creemos que de las tesis defendidas en este artículo se pueden extraer conclusiones interesantes sobre la estrategia defendida por Brown (2004). Según Ludlow (1997),

> Primero, no me parece un resultado trivial que o bien el externismo o bien el auto-conocimiento no se den en el mundo actual. Dicho de otro modo, eso es simplemente decir que una de las doctrinas es falsa, y probar *eso* era primer término la motivación para proponer el argumento incompatibilista.[130]
>
> Segundo, probar que se da este resultado en el mundo actual vendría a probar incluso más de lo pretendido. Entiendo que lo que quería defender Boghossian era que al menos hay un mundo, no importa que sea el mundo actual o no, donde no podemos compatibilizar el externismo y el auto-conocimiento – donde las doctrinas nos llevan a contradicción. Esto es suficiente para mostrar que hay algo horriblemente malo o bien con nuestra concepción del auto-conocimiento o bien con la concepción externista del contenido mental. Demuestra que en ciertas circunstancias concebibles nuestras suposiciones nos llevarán a inferir una contradicción. Pero se supone que los conceptos bien trabajados no deben admitir esto. Si son conceptos de los que se puede hacer buen uso, no deberían llevar a resultados catastróficos en un mundo cercano.[131]

No es aceptable proponer, como respuesta a los argumentos incompatibilistas como los presentados por Boghossian o Brown, que los escenarios de transición lenta no son relevantes (en el caso de Brown, "normalmente relevantes"). Estos argumentos no pretenden demostrar que no hay ningún mundo posible donde las dos tesis son verdaderas, sino plantear escenarios en los que la mera aceptación de una semántica externista nos lleva a concluir que hay individuos en esos escenarios que no pueden conocer mediante introspección qué están pensando. El incompatibilista presupone que ya esta conclusión es indeseable (ni Brown ni Warfield discuten este punto concreto) y

[130] First, it doesn't seem to me a trivial result that either externalism or self-knowledge fail to hold in the actual world. Put another way, that is just to say that one of the doctrines is false, and showing *that* was the motivation for advancing the incompatibilist argument in the first place. (Ludlow (1997), pp. 235-236).

[131] Second, showing that this result holds in the actual world amounts to overkill if anything. I take it that Boghossian's point was that there is at least one world, never mind the actual world, where we cannot square externalism and self-knowledge – where the doctrines lead to contradiction. This is enough to tell us that there is something horribly wrong with either our conception of self-knowledge or the externalist conception of mental content. It shows us that in certain quite conceivable circumstances, our assumptions will lead us to infer a contradiction. But well-honed concepts are not supposed to allow this. If they are serviceable they should not lead to catastrophic results in a nearby world. (236)

concluye que algo no funciona en alguna de las dos tesis si en algún escenario la verdad de la una nos lleva a la falsedad de la otra. Y peor lo tendrá quien quiera responder al argumento simplemente negando la relevancia de las transiciones lentas si, con Ludlow y Brown, acepta que el mundo actual contiene ejemplos de transición lenta.[132]

Resumiendo, creemos que Brown (2004) no propone una buena respuesta al argumento de la discriminación. No es realmente una respuesta al argumento, ya que acepta que un individuo en las condiciones descritas no podría conocer mediante introspección qué está pensando (y esto es lo que en principio el argumento quería probar). Además, se sigue que hay individuos en el mundo actual que no pueden conocer mediante introspección qué están pensando y, nos tememos, el conjunto de individuos en esta situación es mayor de lo que piensa (incluiría a filósofos como la misma Brown). Ese resultado ya nos parece inaceptable. Por otro lado, asume que, en cierto tipo de escenarios, uno no puede saber que está pensando un pensamiento perceptivo *p* basado en ostensión, pero hemos visto que si esto es verdadero, es una verdad que no se puede conocer—y esto nos parece extraño. Para terminar, creemos, siguiendo a Ludlow (1995a, 1997), que las estrategias de este tipo no entienden la dialéctica de la discusión propuesta por el incompatibilista.

[132] También Goldberg (2006) critica esto mismo a Brown: "La principal preocupación detrás de estos problemas no es que el externismo semántico provoca que los fallos de auto-conocimiento sean *demasiado frecuentes*; más bien es que el externismo semántico hace del auto-conocimiento rehén de cambios "externos". Una intuición clave sobre este resultado es que los cambios en el entorno "externo" *nunca* deberían interponerse en la capacidad de un sujeto para obtener conocimiento no-empírico de sus propios pensamientos. Y es justo esta intuición clave la que cede la respuesta de Brown." ("The core worry behind the achievement problem is not that AI makes failures of self-knowledge *too frequent* an occurrence; rather, it is that AI makes self-knowledge hostage to "external" changes in the first place. A key intuition on this score is that changes in the "external" environment should *never* get in the way of a thinker's achieving non-empirical knowledge of her own thoughts. Yet it is precisely this key intuition that Brown's position surrenders." (Goldberg (2006), p. 311))

7. ÚLTIMOS COMENTARIOS Y CONCLUSIONES

Resumamos lo expuesto hasta ahora. Hemos comenzado presentando dos versiones distintas del argumento incompatibilista que queríamos estudiar en esta primera parte, la de Boghossian (1989) y la de Brown (2004). Según la versión de Boghossian, si el modelo observacional del auto-conocimiento y el externismo semántico son ambos verdaderos, la víctima de una transición lenta no podrá conocer qué está pensando mediante introspección, porque su evidencia introspectiva resultará compatible con un escenario alternativo relevante. Según el argumento de la discriminación de Brown, la víctima de una transición lenta no podrá conocer mediante introspección qué está pensando, porque será incapaz de distinguir (mediante introspección) entre su escenario actual y cierto escenario alternativo relevante. Hemos estudiado con detenimiento algunas de las características y presuposiciones de estos argumentos, especialmente aquéllas que tienen que ver con la relación entre los principios (RA) y (Discp) o entre la evidencia que tiene uno y sus capacidades discriminatorias. Hemos sugerido que, dada esta relación entre evidencia y discriminación, no hay grandes diferencias entre los dos argumentos, ya que uno podrá discriminar entre dos escenarios sólo si la evidencia que tiene no es compatible con esos dos escenarios.

Hemos continuando presentando y evaluando algunas de las respuestas que se han ofrecido a estos argumentos. Por ejemplo, hemos visto que alguien podría sentirse

tentado a negar la relevancia de los escenarios alternativos; Brown (2004) argumenta que estos escenarios no son *normalmente* relevantes. Pero hemos defendido que ésta no es una buena opción; primero, porque muerde la bala lanzada por Boghossian y, segundo, porque no entiende bien la dialéctica de la discusión. Además, hemos visto que morder la bala en el caso de los pensamientos perceptivos basados en ostensión y, así, asumir que en algunas situaciones concretas en las que uno tiene un pensamiento de este tipo no sabe qué esta pensando tiene una consecuencia extraña. La consecuencia extraña es que, incluso si la hipótesis de que en esos escenarios no podríamos saber qué estamos pensando fuera verdadera, nadie puede saber que esa hipótesis es verdadera.

En el cuarto capítulo hemos presentado la propuesta de Falvey y Owens (1994), en el quinto la de McLaughlin y Tye (1998); nos hemos basado en sus opiniones para bosquejar la que creemos es la respuesta más adecuada al argumento. Falvey y Owens señalan que (RA) está en la base de los argumentos incompatibilistas como el presentado, y defienden que este principio es falso (algunos les critican que no proporcionan suficientes argumentos en esta dirección, nosotros hemos puesto un contraejemplo para demostrar que, una vez asumimos que las propiedades internas de un sujeto bastan para individuar su evidencia, (RA) y (Discp) son falsos). McLaughlin y Tye responden que el externista puede aferrarse a principios epistémicos como (RA), ya que los argumentos incompatibilistas presuponen una noción de la evidencia introspectiva que no es aceptable.

Basándonos en estas opiniones, hemos propuesto lo que hemos venido a llamar una *disyunción incómoda*: uno ha de elegir entre una noción internista de la evidencia y principios como (RA) y (Discp). Porque es falso que, para toda instancia de conocimiento, ésta tiene que basarse en evidencia que se individúa internamente y que, al mismo tiempo, (RA) o (Discp) imponen condiciones que han de cumplir—la conjunción de (RA) (o (Discp)) y una noción tal de la evidencia tiene contraejemplos. El problema es que los argumentos incompatibilistas típicos presuponen tanto lo uno como lo otro; por eso, el argumento presentado no puede demostrar que el externismo semántico es incompatible con el auto-conocimiento autoritativo.

Ante la disyunción, hemos apostado por rechazar (RA) y (Discp)—seguimos a Falvey y Owens en esto, pues. Oscar no muestra actitudes distintas acerca de esos pensamientos

entre los cuales se supone que puede discriminar, y si descubriera que ha estado transitando de un escenario a otro y pensando sobre cosas distintas, no sabría en qué escenario había estado en cada momento.

Esto sugiere que Oscar no tiene las capacidades discriminatorias que le suponen algunos, y que la tesis de la transparencia del contenido es falsa—sugerimos que el externista debería negar que el contenido es transparente. A lo largo de este trabajo diremos más sobre estas cuestiones; por ahora nos basta con señalar que, aunque fuera verdad que el externismo tiene estas consecuencias (nosotros creemos que lo es), no se seguiría que este modelo muestra algún sesgo de incompatibilidad con la tesis del acceso privilegiado. Del hecho de que S pueda creer erróneamente que dos de sus pensamientos tienen el mismo contenido no se sigue que S puede equivocarse a la hora de identificar el contenido de un pensamiento que está teniendo; para llegar a estas conclusiones incompatibilistas uno necesita de otras asunciones (principios como (RA) o (Discp) y una noción internista de la evidencia, por ejemplo). En contra de lo que parecen pensar algunos, la transparencia del contenido no es una condición necesaria del auto-conocimiento autoritativo.

La estrategia compatibilista que hemos favorecido además va en contra del modelo observacional del auto-conocimiento en la base del argumento. Parece plausible pensar que el conocimiento perceptivo está regido por principios como (RA) o (Discp) y que la evidencia perceptiva se individúa internamente; en cuanto hemos defendido que una de esas dos opciones ha de ser falsa en el caso del auto-conocimiento basado en introspección, se sigue que o bien hay principios epistémicos que rigen el conocimiento perceptivo pero no el auto-conocimiento, o bien que debemos individuar la evidencia de modo distinto—no somos muy originales al afirmar esto: las respuestas de Burge (1988), Falvey y Owens (1994) y McLaughlin y Tye (1998) vienen a rechazar el modelo observacional de auto-conocimiento en la base del argumento.

Sobre estas cuestiones, en un capítulo dedicado a Tyler Burge, hemos aceptado con éste que el conocimiento de los contenidos de nuestros estados mentales se asemeja al conocimiento de proposiciones como “Estoy aquí ahora”. Esto es, como diría Boghossian, este conocimiento se basa en una epistemología insustancial, no es falible, y no hay posibilidad de error “bruto”. Dado que algunos de nuestros juicios sobre

nuestros estados mentales sí son falibles, esto sugiere que en esos casos no fallamos a la hora de identificar el contenido del estado mental en cuestión (que es lo que el externismo semántico podría amenazar), sino a la hora de determinar si guardamos algún tipo de actitud hacia ese contenido (que identificamos correctamente). Dicho de una forma un tanto basta, cuando hacemos un juicio incorrecto sobre en qué estado mental intencional nos encontramos, esto no es así porque "no conocemos contenidos", sino porque no conocemos las relaciones que guardamos hacia esos contenidos.

(2).....

MEMORIA, CAPACIDAD CONCEPTUAL Y AUTO-CONOCIMIENTO

0. INTRODUCCIÓN

En esta segunda parte trataremos cuestiones relacionadas con la memoria. Nos centraremos en la historia de Sally, una chica que, al igual que Oscar, viaja inadvertidamente de un entorno de agua a otro de bi-agua. Al igual que Oscar, Sally pasa su infancia en la Tierra, rodeada de agua, H_2O—Sally bebe agua, hierve las patatas en agua, se ducha con agua, habla del agua (profiriendo enunciados que contienen el término 'agua') y piensa y guarda creencias acerca del agua (pensando pensamientos y creyendo creencias en parte constituidos por el concepto AGUA). Supongamos ahora que, un día, cuando todavía habita la Tierra y no ha sufrido ningún tipo de transición a la Tierra Gemela, en t1, la pequeña Sally visita por primera vez la costa. Bañándose en el mar por primera vez, piensa un pensamiento que expresaría profiriendo el siguiente enunciado:

t1: El agua es a veces salada.

Bien, hasta ahí nada raro. Pero sucede que un día años más tarde, al igual que Oscar, Sally cambia inadvertidamente de escenario; a pesar de que se va a la cama en la Tierra, amanece en la Tierra Gemela. Como sabemos, la Tierra y la Tierra Gemela son muy parecidas: hay humanos que habitan la Tierra Gemela, que llevan un estilo de vida parecido al que llevan los humanos en la Tierra a comienzos del siglo XXI, viven en

ciudades, viajan en coche, adoptan perros y gatos como animales de compañía, y pasan el tiempo viendo programas de telerrealidad. De hecho, la Tierra y la Tierra Gemela se parecen tanto entre sí que Sally nunca se percata de que ha cambiado de la una a la otra. La única diferencia entre la Tierra y la Tierra Gemela radica en que ésta no contiene agua, H_2O, sino una sustancia que llamamos 'bi-agua', y cuya composición química es XYZ. El agua y la bi-agua tienen las mismas propiedades *macro*, y los habitantes de la Tierra Gemela usan el término 'agua' para referirse a la bi-agua. Sally vive en la Tierra Gemela el tiempo suficiente para adquirir el concepto BI-AGUA y poder referirse a la bi-agua cuando profiere enunciados que contienen el término 'agua'. A pesar de ello, en ningún momento descubre que ha habitado dos escenarios distintos, y tampoco aprende que la bi-agua es XYZ (nunca supo que el agua es H_2O).

Un día, en la Tierra Gemela, Sally comienza a recordar su primera visita a la costa cuando todavía era niña—recuerda un pensamiento que tuvo cuando se bañó en el mar por primera vez. En t2, Sally tiene un pensamiento que expresaría si profiriera el siguiente enunciado:

t2: En t1 pensé que el agua es a veces salada.

Fin de la historia. Basándose en casos como el de Sally, Boghossian (1989) propuso un argumento incompatibilista, "El Argumento de la Memoria". Viene a decir que, si el externismo semántico tiene como consecuencia que sujetos como Sally tienen dificultades para recordar antiguos pensamientos que tuvieron, entonces el externismo resulta incompatible con el auto-conocimiento autoritativo.

En esta segunda parte presentaremos el argumento propuesto por Boghossian primero, y describiremos y discutiremos después las diferentes respuestas que tiene a mano el compatibilista. Por el camino nos centraremos en cuestiones relativas a las capacidades mnemónicas y conceptuales que tendría alguien en la situación de Sally; volveremos varias veces al escenario que acabamos de presentar, en un capítulo discutiremos también una variación de este escenario, e incluso presentaremos lo que llamamos "una (proto)teoría de la memoria".

Pero comencemos por presentar el argumento tal y como lo expuso Boghossian.

1. EL ARGUMENTO DE LA MEMORIA

Es en las últimas páginas de "Content and Self-Knowledge" donde Boghossian presenta lo que se ha venido a llamar "el argumento de la memoria". Su propósito es demostrar que, incluso si dejáramos de lado las principales críticas a la estrategia compatibilista de Burge, dados ciertos compromisos del externismo semántico sobre la memoria, los pensamientos *cogito* no pueden constituir conocimiento.

Recordemos primero cuáles eran los ejes de la crítica de Boghossian a la estrategia compatibilista de Burge:

- El auto-conocimiento es falible (a diferencia del conocimiento de proposiciones empíricas basado en epistemología insustancial).
- Los pensamientos *cogito* como mucho podrían explicar una parte mínima de nuestro auto-conocimiento. Los pensamientos *cogito* no son auto-verificantes si son sobre:
    - Estados mentales *permanentes*.
    - Estados mentales intencionales compuestos por una actitud proposicional que no sea la de "pensar".
    - Estados mentales pasados.

Pero, independientemente de que no poder explicar casos paradigmáticos del auto-conocimiento suponga (o no) un grave problema para la estrategia de Burge, el mismo Boghossian acepta que éste sí ofrece algún tipo de respuesta a su argumento incompatibilista si los pensamientos *cogito* que describe nos proveen de algún tipo de conocimiento:

> Aun así, a pesar de que la propuesta de Burge no explica los casos centrales, ¿no nos provee al menos *un* caso en el que un pensamiento es conocido directamente a pesar de la naturaleza relacional de sus condiciones de individuación? ¿Y no es eso suficiente para echar por tierra nuestra intuición de que el relacionismo es irreconciliable con el conocimiento directo de nuestros pensamientos? Si los juicios auto-verificantes de Burge fueran instancias de conocimiento genuino, de seguro echarían por tierra la problemática intuición.[133]

El argumento de la memoria intenta probar que ni siquiera los pensamientos *cogito* nos aportan auto-conocimiento de ningún tipo. Boghossian dedica únicamente un par de párrafos al argumento, y por eso los transcribimos aquí casi íntegramente:

> Burge observa:
>
> > (...) Por supuesto, la persona víctima de una transición lenta podría aprender acerca de las transiciones y preguntar "¿Estaba yo ayer pensando sobre agua o bi-agua?" – y de hecho no conocer la respuesta. Aquí conocer la respuesta puede algunas veces depender de conocer algunas condiciones empíricas de fondo. Pero tales cuestiones sofisticadas sobre la memoria requieren una historia más compleja. (Burge (1988), p. 659)
>
> Estos apuntes me resultan realmente chocantes. Vienen a decir que, a pesar de que S no sabrá mañana qué está pensando ahora mismo, sí sabe ahora mismo qué está pensando ahora mismo. Para cualquier momento en el presente, digamos que t1, S está en la posición de pensar un juicio auto-verificante sobre qué está pensando en t1. Según los criterios de Burge, por lo tanto, en t1 tiene conocimiento directo y autoritativo de qué está pensando en ese momento. Pero está más bien claro que mañana no sabrá qué pensó en t1. En ese momento no tendrá a mano ningún juicio auto-verificante sobre su pensamiento en t1. Para saber qué pensó en t1 debe descubrir en qué entorno estaba en aquel momento y cuánto tiempo estuvo en él. Pero hay un misterio aquí. Porque lo siguiente viene a parecer una obviedad sobre memoria y conocimiento: Si S sabe que p en t1, y si (más tarde) en t2, S recuerda todo lo que S sabía en t1, entonces S sabe que p en t2. Ahora, preguntémonos: ¿*por qué* no sabe S hoy si el pensamiento de ayer era un pensamiento *de agua* o un pensamiento *de bi-agua*? La verdad obvia insiste en que sólo hay dos explicaciones posibles: o bien S ha olvidado o bien *nunca* supo. Pero es seguro que un fallo de la memoria no es el caso aquí. Cuando discutimos la epistemología de contenidos individuados relacionalmente, deberíamos poder excluir fallos de la memoria

[133] Still, even if Burge's proposal does not explain the central cases, does it not supply us with at least *one* case in which a thought is known directly despite the relational nature of its individuation conditions? And isn't that enough to dislodge our intuition that relationism is irreconcilable with directness? If Burge's self-verifying judgments were instances of genuine knowledge, then they would indeed dislodge the problematic intuition. (Boghossian (1989), p. 171)

> mediante estipulación. Parece que los pensamientos con contenidos individuados ampliamente no son fáciles de conocer, sino difíciles de recordar. La única explicación, me atrevería a sugerir, de por qué S no sabrá mañana lo que dicen que sabe hoy, no es que ha olvidado, sino que nunca supo. Los juicios auto-verificantes de Burge no constituyen conocimiento genuino.[134]

En t1, Sally juzga "Estoy pensando que el agua es a veces salada". Este juicio de segundo orden está en parte constituido por el pensamiento de primer orden acerca del cual es y, por eso mismo, es auto-verificante. Pero sucede que en t2, cuando está en un entorno que no contiene agua sino bi-agua, Sally no puede tener un juicio auto-verificante sobre aquello que pensó en t1. Parece que, de este hecho y de la breve cita de Burge (1988), Boghossian concluye que si el externismo semántico es verdadero, Sally no puede saber en t2 mediante introspección qué pensaba en t1. Pero hay un problema aquí, nos dice Boghossian, ya que entiende que es una *obviedad* acerca de la memoria que, si S no sabe que *p* en t2 entonces, o bien S nunca supo que *p*, o bien S ha olvidado que *p*—y podemos estipular que Sally nunca olvidó nada. Así, concluye Boghossian, se sigue que Sally no puede saber en t1 mediante introspección que está pensando que el agua es a veces salada, independientemente de los pensamientos *cogito* que pueda tener; esos pensamientos *cogito* podrán ser veraces, auto-verificantes, pero no constituyen conocimiento. Por lo tanto, la estrategia compatibilista de Burge falla—el externismo semántico y el auto-conocimiento autoritativo son incompatibles.

---

[134] Burge observes:

> (...) Of course, the person [victim of a slow switch] may learn about the switches and ask "Was I thinking yesterday about water or twater?" – and yet not know the answer. Here knowing the answer may sometimes depend on knowing empirical background conditions. But such sophisticated questions about memory require a more complex story. (Burge (1988), p. 659)

These remarks strike me as puzzling. They amount to saying that, although S will not know tomorrow what he is thinking right now, he does know right now what he is thinking right now. For any given moment in the present, say t1, S is in a position to think a self-verifying judgment about what he is thinking at t1. By Burge's criteria, therefore, he counts as having direct and authoritative knowledge at t1 of what he is thinking at that time. But it is quite clear that tomorrow he won't know what he thought at t1. No self-verifying judgment concerning his thought at t1 will be available to him then. To know what he thought at t1 he must discover what environment he was in at that time and how long he had been there. But there is a mystery here. For the following would appear to be a platitude about memory and knowledge: If S knows that p at t1, and if at (some later) t2, S remembers everything S knew at t1, then S knows that p at t2. Now let us ask: *why* does S not know today whether yesterday's thought was a *water* thought or a *twater* thought? The platitude insists that there are only two possible explanations: either S has forgotten or he *never* knew. But surely memory failure is not the point. In discussing the epistemology of relationally individuated content, we ought to be able to exclude memory failure by stipulation. It is not as if thoughts with widely individuated contents might be easily known but difficult to remember. The only explanation, I venture to suggest, for why S will not know tomorrow what he is said to know today, is not that he has forgotten but that he never knew. Burge's self-verifying judgments do not constitute genuine knowledge. (Boghossian (1989), pp. 171-172)

Ludlow (1995b) esquematiza muy adecuadamente el argumento:

(1) Si S no olvida nada, entonces lo que sabe S en t_1, sabe S en t_2,
(2) S no olvidó nada,
(3) S no sabe que P en t_2;
(4) Por lo tanto, S no sabía que P en t_1.[135]

Muy brevemente, el argumento principalmente asume dos tesis. Por un lado, asume que el externista está comprometido a aceptar que alguien en la situación de Sally no podría conocer mediante introspección, en t2, qué estaba pensando en t1. Por otro lado, también se basa en lo que Boghossian llama una *obviedad* sobre la memoria y el conocimiento; que si alguien no sabe que *p* en t2, eso sólo tiene dos explicaciones posibles: o bien S nunca supo que *p* o bien S ha olvidado que *p*.

Nos gustaría mencionar un par de cuestiones acerca del alcance y el objetivo del argumento. Dijimos, sobre el argumento presentado en la primera parte de este trabajo, que su conclusión era que un individuo en la situación de Oscar no podría conocer (sólo mediante introspección) qué estaba pensando; era en principio una cuestión abierta si el argumento podía demostrar que el externismo semántico amenaza de algún modo *nuestro* auto-conocimiento autoritativo. La misma duda surge con el argumento de la memoria: quizás sólo amenace el auto-conocimiento autoritativo de Sally. Ya defendimos que el primer argumento sí amenazaba nuestro propio auto-conocimiento (sección 6.4. en la primera parte), y este segundo argumento se generaliza aún más fácilmente. Su conclusión es que en t1 Sally no sabe qué está pensando. Pero en t1 Sally todavía no ha sido víctima de una transición lenta, y no parece asumible que sean las futuras contingencias de sus aventuras las que minen su auto-conocimiento en t1—por eso, una vez aceptamos con Ludlow (1995a) que en el mundo actual sí se pueden dar transiciones lentas de este tipo, se sigue que la conclusión del argumento de la memoria es que *nuestro* auto-conocimiento autoritativo es incompatible con el externismo semántico.[136]

[135] (1) If S forgets nothing, then what S knows at t_1, S knows at t_2, (2) S forgot nothing, (3) S does not know that P at t_2; (4) therefore, S did not know that P at t_1. (Ludlow (1995b), p. 308)
[136] Brueckner (1997): "Supongamos que Boghossian puede demostrar que en esas situaciones, un individuo que está en la Tierra no sabe en t, vía el mecanismo de Burge, qué está pensando en t, en virtud de las transiciones *subsiguientes* y la aplicación del razonamiento sobre memoria. Sería extremadamente

A diferencia del argumento incompatibilista estudiado en la primera parte, no parece que este segundo argumento se base en un modelo observacional del auto-conocimiento. Pero sí es una premisa del argumento que alguien como Sally no podría saber, en t2 y sólo mediante introspección, qué pensó en t1. Boghossian no da demasiadas razones a favor de esta tesis. Cita un párrafo de Burge (1988) donde éste dice que, si supiera sobre su historial de transiciones lentas, Sally no podría responder a la pregunta '¿en t1 pensabas sobre agua o sobre bi-agua?'. Creemos que el hecho de que el sujeto sepa acerca de sus historial de transiciones es un factor decisivo en el diagnóstico que hace Burge (en el tercer y cuarto capítulo pondremos más atención a estas cuestiones). La cita no dice nada acerca de las capacidades mnemónicas de Sally en t2 y, además, en "Memory and Self-Knowledge" (1998) Burge dice claramente que con ese párrafo no pretendía insinuar que la víctima de una transición lenta no podría recordar sus pensamientos, que se refería a una pregunta concreta hecha cuando el individuo tiene una información concreta. Por todo lo que dice Burge (1988), es una cuestión abierta si en t2 Sally puede recordar qué pensó en t1.

Pero el argumento de la memoria no llega a demostrar que el externismo semántico amenaza nuestro auto-conocimiento autoritativo. Una de sus premisas ha de ser necesariamente falsa. Así lo defenderemos en el siguiente capítulo (además, explicaremos por qué el argumento presenta una cuestión interesante sobre externismo y memoria).

---

extraño suponer que el mecanismo de auto-verificación es *alguna vez* suficiente para asegurar el conocimiento de qué está pensando uno cuando lo está pensando." ("Suppose that Boghossian can show that in those situations, a thinker who starts out on Earth does not know at t, via Burge's mechanism, what he is thinking at t, in virtue of the *subsequent* switches and the application of the reasoning about memory. It would then be extremely odd to suppose that the mechanism of self-verification is *ever* sufficient to secure knowledge of what one is thinking when one is thinking it." (Brueckner (1997), p. 322))

2. TRANSICIONES LENTAS, CAPACIDAD CONCEPTUAL

Una posible respuesta al argumento de la memoria se reduce a simplemente rechazar la premisa (3), que asume que la víctima de una transición lenta no podría saber qué pensó en el escenario anterior.

El ejemplo que discutimos en esta parte nos presenta a Sally, víctima de una transición lenta, quien en t1 piensa que el agua es a veces salada y en t2, una vez que está en la Tierra Gemela y ha adquirido el concepto BI-AGUA, intenta recordar qué pensó en t1. En t2 Sally está en una situación mental que expresaría profiriendo el enunciado 'en t1 pensé que el agua es a veces salada'; pero, ¿cuál es, según el externista, el contenido de ese pensamiento? De acuerdo con el externismo de clases naturales, qué clases se encuentran presentes en nuestro entorno es un factor relevante a la hora de individuar el contenido de nuestros pensamientos. Ahora, parece que en el ejemplo descrito hay dos entornos distintos que, en principio, son buenos candidatos a ser el entorno relevante a la hora de individuar el contenido del pensamiento de Sally: su entorno en t1, que contiene agua, donde piensa el pensamiento que trata de recordar en t2; y su entorno en t2, que contiene bi-agua, donde intenta recordar el pensamiento que pensó en t1. Dependiendo de qué entorno crea el externista que es el relevante, defenderá que el pensamiento de Sally en t2 tiene un contenido u otro.

2.1. COHABITACIÓN DE CONCEPTOS: SALLY SÍ RECUERDA QUÉ PENSÓ

Algunos autores[137] defienden que es el entorno de Sally en t1 el relevante para determinar el contenido de su pensamiento en t2 y que, por eso, Sally no pierde su concepto AGUA cuando adquiere el concepto BI-AGUA. Esto es, según estos autores las transiciones lentas son ejemplos de *cohabitación de conceptos* y no de *reemplazo conceptual*. Es probable que sea Burge (1998) el ejemplo más notable de una propuesta de cohabitación de conceptos. Según él, "mudarse al otro entorno y adquirir nuevos conceptos no elimina los conceptos y memorias antiguos que devienen del primer entorno"[138], "no [ve] ninguna razón para pensar que S normalmente perderá el concepto original"[139]. Parece que, de acuerdo con Burge, simplemente no hay motivos suficientes para pensar que Sally carece del concepto AGUA en t2.

Burge distingue entre lo que denomina 'memoria preservativa' y 'memoria sustancial' y mantiene que, mientras se base en la memoria preservativa, Sally no tendrá ningún problema para recordar en t2 qué pensó en t1. Citando a Burge,

> La memoria preservativa normalmente mantiene el contenido y los compromisos de actitud de pensamientos anteriores, mediante conexiones causales a los pensamientos pasados. Ésa es una de sus funciones – mantener y preservar un punto de vista a lo largo del tiempo. No necesita tomar un pensamiento pasado como objeto de investigación, en necesidad de discriminación de otros pensamientos.[140]

Esto es, muy brevemente, según Burge hay un proceso cognitivo que se llama 'memoria preservativa', la cual es distinta a la "memoria sustancial". La primera diferencia entre los dos tipos de memoria es que cumplen funciones distintas. La memoria preservativa "le trae a la mente" al sujeto, vía relación causal, pensamientos que ya ha tenido antes.

[137] Burge (1998), Gibbons (1996), Kraay (2002), Schiffer (1992).
[138] Moving to the other environment and acquiring new concepts will not normally obliterate old concepts or memories that derive from the first environment. (Burge (1998), p. 356)
[139] I see no reason to think that S will ordinarily lose the original concept (Burge (1998), p. 369)
[140] Preservative memory normally retains the content and attitude commitments of earlier thinkings, through causal connections to the past thinkings. That is one of its functions – maintaining and preserving a point of view over time. It need not take a past thought as an object of investigation, in need of discrimination from other thoughts. (Burge (1998), p. 357)

La memoria preservativa nos posibilita recordar nuestras creencias, nos permite "mantener un mismo punto de vista a lo largo del tiempo". Gracias a la memoria preservativa podemos recordar, por ejemplo, que Platón escribió *La República*, que tenemos que coger la línea amarilla del metro para llegar a casa, o que los libros de filosofía son generalmente aburridos. Por contra, la memoria sustancial nos permite recordar hechos y eventos de nuestra autobiografía; este tipo de memoria permite a Platón recordar que anoche terminó de escribir *La República*, a Maider que la vez que cogió la línea azul del metro terminó en Cornellá, o a Rosa que el día que leyó *Ser y Tiempo* se aburrió como nunca antes.

Pero según Burge, además, la memoria preservativa y la sustancial funcionan de diferente manera. La memoria sustancial *identifica* el episodio pasado mediante referencias a objetos, actitudes, imágenes o eventos, mientras que la preservativa "trae a la mente" recuerdos gracias a la mediación de cadenas causales. Es gracias a estas cadenas causales que el pensamiento recordado adquiere el contenido del pensamiento original: la memoria "capta" el antecedente automáticamente, sin necesidad de identificación alguna[141]. Así, aunque haya una transición lenta de por medio, si Sally recuerda un pensamiento mediante su memoria preservativa, no hay peligro de que "confunda" el contenido de ese pensamiento con otro. En cuanto de hecho se den estas cadenas causales entre el pensamiento original y su recuerdo, el recuerdo guardará el contenido original del pensamiento. Además, el sujeto estará justificado a la hora de apoyarse en la memoria preservativa para recuperar el conocimiento que una vez tuvo; la memoria preservativa ni añade ni resta nada a la justificación de la creencia original. Por contra, la memoria sustancial refiere al episodio pasado mediante *identificación*, y no se basa en una cadena causal que preserve ningún contenido original. Así, la memoria sustancial no es inmune a las transiciones lentas, es no-preservativa—Sally en principio puede confundir un episodio de agua por uno de bi-agua.

Pero centrémonos en la memoria preservativa, en qué recuerdos de agua puede tener Sally en t2, después de completarse la transición lenta. Si recuerda el pensamiento que

[141] Burge propone cierta analogía entre la memoria preservativa y la anáfora: el hablante o el oyente no necesitan *identificar* la referencia de las expresiones anafóricas para que esta expresión pueda referir correctamente, basta con que la expresión anafórica en cuestión se base en cadenas referenciales inherentes al discurso; cuando alguien recuerda un pensamiento mediante la memoria preservativa, basta con que las cadenas causales en las que se basa la memoria preservativa *se den*.

pensó en t1, que el agua es a veces salada, mediante la memoria preservativa, entonces aquello que recuerda preserva el contenido de aquel pensamiento original, y dado que se basa en este mecanismo, Sally puede recordar y saber que el agua es a veces salada (aun y cuando no es capaz de distinguir este pensamiento del pensamiento de que la bi-agua es a veces salada). Pero una cosa es recordar que el agua es a veces salada y otra recordar que en un momento dado se pensó que el agua es a veces salada. El primero es un recuerdo de un pensamiento que se ha tenido antes y, por lo tanto, es la memoria preservativa la encargada de proveerle este recuerdo al sujeto; el segundo recuerdo, en cambio, es el recuerdo de un episodio vivido por Sally y, en principio, es la memoria sustancial (no-preservativa) la encargada de proporcionarle este conocimiento. Así, en un principio parecería que Sally no puede saber en t2 que en t1 pensó que el agua es a veces salada, parece que podría identificar erróneamente ese pensamiento como que en t1 pensó que la bi-agua es a veces salada.

A pesar de ello, Sally puede en t2 saber que en t1 pensó que el agua es a veces salada según Burge si la memoria preservativa y la sustancial "trabajan juntas":

> El individuo podría recordar el pensamiento pasado como un evento; y sólo el contenido del pensamiento pasado podría ser recordado en el modo preservativo. Su memoria enlazaría luego el uso conceptual actual a los conceptos de ese pensamiento pasado. Aquí la memoria sustantiva de eventos y la memoria preservativa trabajarían juntas – la primera identificando un evento, la segunda preservando el contenido y la modalidad de actitud del evento.[142]

Esto es, la memoria preservativa y la sustancial habrán trabajado juntas si Sally recuerda en t2 que en t1 pensó que el agua es a veces salada. La memoria sustancial identifica el evento y el momento en cuestión: el acto de pensar que el agua es a veces salada y t1. Sally recuerda el pensamiento de que el agua es a veces salada mediante la memoria preservativa (lo cual posibilita que su pensamiento en t2 herede el concepto AGUA que constituía el pensamiento original en t1), todos los demás elementos mediante la memoria sustancial, por lo que en t2 puede saber que en t1 pensó que el agua es a veces salada.

[142] The individual could remember the past thinking as an event; and only the content of the thinking could be remembered in the preservative way. His memory would then tie current conceptual use to the concepts of that past thinking. Here substantive event memory and preservative memory would work together – the former identifying an event, the latter preserving the content and attitude-modality of the event. (Burge (1998), p. 359)

No creemos que la propuesta de Burge sea convincente. No está claro cómo es eso de que la memoria sustancial y la preservativa "trabajen juntas", y no vemos motivo alguno para diferenciar entre memorias preservativas y memorias no-preservativas. Así lo defenderemos en el capítulo 6 de esta parte.

Gibbons (1996) también apuesta por la cohabitación de conceptos. Defiende una interpretación causal del auto-conocimiento[143]; según él, cuando un pensamiento de primer orden *causa* un pensamiento de segundo orden, éste hereda los conceptos y contenidos del primero (debido a esa relación causal). Adoptando, como hace Gibbons, una teoría fiabilista en epistemología, además, no parece que haya demasiados problemas a la hora de defender que podemos *saber* qué estamos pensando. Ya hemos dedicado toda la primera parte de este trabajo a estas cuestiones, pasemos a temas relacionados con la memoria.

Gibbons no ve ningún motivo para pensar que en las transiciones lentas se da reemplazo conceptual: "Siendo fácil ver cómo un contacto causal con un nuevo tipo de sustancia puede proveerte con un nuevo concepto, no está del todo claro cómo te puede privar de uno"[144]. Así, una vez suponemos que Sally puede tener en t2 los dos conceptos AGUA y BI-AGUA y abogamos, junto con Gibbons, por una interpretación causal del auto-conocimiento, parece fácil explicar cómo puede Sally saber en t2 qué pensó en t1:

> El mero hecho de transitar a la Tierra Gemela no cambia el contenido de lo que está guardado en la memoria. El contenido de lo que está guardado está determinado por la historia causal del estado guardado. El contenido de un estado que es inferido (conscientemente o no) exclusivamente de estados guardados está determinado por el contenido de los estados guardados. El contenido de un estado inferido en parte de estados guardados y en parte de información actual está determinado por el contenido de ambos tipos de estados, etc.[145]

Sally piensa conscientemente en t1 que el agua es a veces salada. Este pensamiento, este *estado*, queda guardado en su memoria, adquiriendo así la creencia disposicional de que

[143] En contra de lo que manifiesta el mismo Gibbons, no creemos que sus ideas al respecto se diferencien demasiado de las propuestas de Falvey y Owens (1994).
[144] While it is easy to see how causal contact with a new type of substance can give you a new concept, it is not at all clear how it can take one away. (Gibbons (1996), p. 295)
[145] Merely switching to Twin Earth does not change the content of what is stored in memory. The content of what is stored is determined by the causal history of the stored state. The content of a state that is inferred (consciously or otherwise) exclusively from stored states is determined by the content of the stored states. The content of a state inferred partly from stored states and partly from current information is determined by the content of both sorts of states, and so on. (Gibbons (1996), p. 304)

en t1 pensó que el agua es a veces salada. Como el transitar a la Tierra Gemela no le priva de ningún concepto que tenía antes, el pensamiento guardado y la creencia disposicional mantienen su contenido original. La creencia de segundo orden que tiene Sally en t2 es, por lo tanto, que en t1 pensó que el agua es a veces salada y, dado que si el pensamiento de primer orden hubiera tenido un contenido distinto entonces también habría tenido una creencia de segundo orden con contenido distinto, en t2 Sally recuerda y sabe que en t1 pensó que el agua es a veces salada.

En resumen, Burge (1998) y Gibbons (1996) defienden que no hay razones suficientes para pensar que en t2 Sally carece del concepto AGUA. Además, dadas ciertas relaciones causales, podrá recordar y saber qué pensó en t1. Por eso, la premisa (3) del argumento de la memoria es falsa y, así, no se sigue que en t1 Sally no sabía qué estaba pensando. El externismo semántico no amenaza el auto-conocimiento autoritativo[146].

2.2. REEMPLAZO CONCEPTUAL: SALLY NO RECUERDA QUÉ PENSÓ

Seguramente los defensores del reemplazo conceptual sean mayoría[147]. Peter Ludlow (1995b, 1996, 1999) es un caso paradigmático de defensor de esta idea. Según él,

> El externista social coherente está atado a decir que el contenido de una memoria se fija en el momento en el que el acto de recordar acontece – pues son las circunstancias que envuelven esa memoria las que resultan cruciales para fijar su contenido. La idea de que hay algún contenido que se determina en algún momento inicial y luego se mantiene congelado hasta un momento posterior en el que es recordado parece equivocada. Al fin y al cabo, ¿qué querría decir que los contenidos de nuestras memorias se fijan por nuestro entorno social si de hecho esos contenidos, una vez fijados, resultan totalmente inertes a cualquier cambio en el entorno?[148]

[146] Burge (1998), además, defiende que también las premisas (1) y (2) del argumento de la memoria son falsas. Contra (2), argumenta que podría haber casos en los que la víctima de la transición lenta perdiera el concepto AGUA (así sucedería, por ejemplo, si aprendiera algo que expresaría profiriendo el enunciado 'el agua es XYZ'); en estos casos de reemplazo en los que el sujeto no puede acceder a los contenidos anteriores a la transición, simplemente es falso que no haya olvidado nada (en contra de lo que piensa Boghossian, no podemos *estipular* que Sally no olvida nada). Contra (1), defiende que parece obvio que hay casos en los que alguien sabe que *p* en t1, no ha olvidado nada, pero no sabe que *p* en t2 (porque ha adquirido evidencia que mina la justificación que tenía en t1 a favor de *p*, por ejemplo).

[147] Entre otros, defienden que los casos de transición lenta son ejemplos de reemplazo conceptual: Bernecker (1998), Brueckner (1997), Falvey (2003), Ludlow (1995b, 1996, 1999) y Tye (1998).

[148] The consistent social externalist is bound to say that the content of a memory is fixed at the time recollection takes place – for it is the embedding circumstances of that memory which are crucial to the

Muy brevemente: es el entorno de Sally en t2 el único relevante para determinar qué conceptos tiene en t2 y cuáles son los contenidos de sus estados mentales en t2. Dado que el entorno de Sally en t2 contiene bi-agua y carece de agua, ésta tiene el concepto BI-AGUA pero no el concepto AGUA. Como carece del concepto AGUA, en t2 Sally no puede pensar pensamientos en parte constituidos por ese concepto y, por eso, no puede pensar, recordar o saber que el agua es a veces salada o que en t1 pensó que el agua es a veces salada. Sally erróneamente creerá recordar que en t1 pensó que la bi-agua es a veces salada. La premisa (3) en el argumento de Boghossian es verdadera.

No llegamos a entender por qué el externista coherente está "atado a decir que el contenido de una memoria se fija en el momento en el que el recuerdo acontece", o por qué "la idea de que hay algún contenido que se determina en algún momento inicial parece equivocada"—a nosotros, simplemente, no nos "parece equivocada". Ludlow no ofrece ningún argumento para defender que los casos de transición lenta son casos de reemplazo conceptual; tenemos la impresión de que esta falta de argumentación es bastante común entre los defensores del reemplazo[149].

Ludlow asume, pues, que Sally adquiere el concepto BI-AGUA y pierde el concepto AGUA cuando ya ha pasado el tiempo suficiente en la Tierra Gemela; otros defienden que el reemplazo conceptual es de otro tipo. Según éstos, la víctima de una transición perdería su anterior concepto, pero no lo reemplazaría por el concepto que es común en la nueva comunidad lingüística que habita. Al contrario, adquiriría un *concepto amalgama*. Falvey (2003) propone el ejemplo de Gloria, víctima de una transición lenta, quien en la Tierra Gemela recuerda un día (en la Tierra) en el que fue a la playa.

> La proferencia post-transición de Gloria de 'el agua estaba agitada aquel día en la playa'. Dada su confusión entre agua y bi-agua, he dicho, no podemos simplemente interpretarla como expresando la proposición de que el *agua* estaba agitada aquel día en la playa. Pero aquí es natural decir que su término 'agua' expresa ahora un concepto disyuntivo—llamémoslo *dosaguas*—donde la dosaguas es o bien agua o bien bi-agua. Esto nos permite decir que la proposición que expresa sobre aquel día

---

fixing of its content. The idea that there be some content which is determined at some initial time and then remains frozen up to some later moment of recollection seems wrongheaded. After all, what would it mean to say that the contents of our memories are fixed by our social environment if in fact those contents, once fixed, are totally inert to all environmental changes? (Ludlow (1995b), pp. 308-309)

[149] Tye (1998) sí que presenta un argumento. Dado que éste tiene que ver con las aptitudes lógicas que tendría la víctima de una transición lenta, lo trataremos en la tercera parte.

> en la playa es verdadera, y quizás incluso cuenta como conocimiento; siendo agua, la sustancia en la playa aquel día también era dosaguas.[150]

Según Falvey, la víctima de una transición lenta adquiere un *concepto amalgama*, al estilo del concepto JADE[151]. Los conceptos amalgama no refieren a una única clase natural, sino a un conjunto de clases naturales. Así, el concepto que expresa Sally al proferir el término 'agua' una vez se ha completado la transición es DOSAGUAS, esto es, o bien agua o bien bi-agua. Cuando Sally por ejemplo profiere el enunciado 'el agua es refrescante' lo que está diciendo es que el agua y la bi-agua son refrescantes. Cuando en la Tierra Gemela Gloria profiere el enunciado 'el agua estaba agitada aquel día en la playa' está expresando la proposición de que algo que o bien era agua o bien era bi-agua estaba agitada aquel día en la playa; por lo tanto, lo que recuerda es verdadero.

Pero de esto no se sigue que en t2 Sally pueda recordar qué pensó en t1:

> Por otro lado, no sólo falla en recordar que pensó que el agua estaba agitada, su proferencia de segundo orden 'pensé que el agua estaba agitada' expresa ahora una aparente memoria falsa de haber tenido el pensamiento de que la dosaguas estaba agitada—el concepto *dosaguas* no figuraba en su pensamiento original.[152]

Esto es, en t2, cuando Sally profiere el enunciado 'en t1 pensé que el agua es a veces salada', expresa la proposición de que en t1 pensó que la dosaguas es a veces salada, lo

---

[150] ...Gloria's post-switch utterances of "The water was choppy that day at the beach." Given her conflation of water and twater, I have argued, we cannot straightforwardly interpret her as expressing the proposition that the *water* was choppy that day at the beach. But here it is natural to say that her word 'water' now expresses a disjunctive concept—call it *zwater*—where zwater is either water or twater. This permits us to say that the proposition she expresses about that day at the beach is true, and perhaps even counts as knowledge; being water, the stuff at the beach that day was also zwater. (Falvey (2003), pp. 228-229)

[151] Heal (1998) parece apuntar también en esta dirección (aunque no es del todo explícita al respecto). Heal (1998) se limita al externismo de clases naturales, dejando fuera el externismo social del tipo de Burge, y asume que para todo término de clase natural, hay un conjunto de ítems de esa clase natural que, al mismo tiempo, fijan, tanto el *estándar* de la clase como la referencia del término. Es de suponer que Sally mencionaría instancias de agua y de bi-agua como ejemplos de aquello que llama 'agua' en t2, independientemente de que el término 'agua' se use para expresar un recuerdo o como refiriéndose a una muestra de bi-agua que tiene enfrente. Por eso, si para cualquier uso de 'agua' que haga Sally son parte del *estándar* tanto instancias de agua como de bi-agua, parece que Sally expresará algo parecido a un concepto amalgama cuando profiera enunciados que contengan el término 'agua'. Creemos que Heal (1998) al menos comete un error: ceñirse al externismo de clases naturales. Si el externismo social es verdadero, no podemos asumir que la única diferencia entre la Tierra y la Tierra Gemela es que contienen dos clases naturales diferentes, el que Sally cambie de comunidad lingüística es un factor que no se puede obviar. La propuesta de Heal sería más plausible si Sally transitara a una Tierra Gemela que contuviera bi-agua y que careciera de humanos (y de comunidades lingüísticas por lo tanto).

[152] On the other hand, she not only fails to remember that she thought that the water was choppy, her second-order utterance, "I thought that the water was choppy" now expresses a false apparent memory of having thought that the zwater was choppy—the concept *zwater* did not figure in her original thought. (Falvey (2003), p. 229)

cual es falso (en t1 Sally no tenía el concepto DOSAGUAS). Debido al reemplazo conceptual que sufre, en t2 Sally no puede saber que en t1 pensó que el agua es a veces salada—la premisa (3) del argumento de la memoria es verdadera.

Por supuesto, el externista que defiende que los casos de transición lenta son ejemplos de reemplazo conceptual puede rechazar alguna de las otras dos premisas en el argumento de la memoria. Por ejemplo, Ludlow (1995b) argumenta que la premisa (1) es falsa:

> Boghossian está en lo cierto al aseverar que Sally no sabe en t2 lo que sabía en t1, pero está equivocado en suponer que "la única explicación" para esto es que Sally "nunca conoció" sus pensamientos en el primer momento. Es totalmente coherente con la visión externista de la memoria que Sally no olvidó nada, pero que aun así los contenidos de sus memorias han cambiado. Es más, esto no es sólo *posible* de acuerdo con el externismo social, sino que dado el predominio de las transiciones lentas debería ser un hecho bastante común.[153]

Esto es, para Ludlow, la premisa (2) en el argumento de la memoria, que Sally no olvidó nada, es verdadera. Parece que, con Boghossian, acepta que al describir el caso de transición lenta simplemente podemos estipular que eso fue así. Pero la premisa (1), que si Sally no olvidó nada, entonces si en t1 sabía que *p* entonces en t2 sabe que *p*, sí es falsa según Ludlow. Uno puede saber que *p* en t1, no olvidar nada, y no saber que *p* en t2; de hecho, el caso de Sally es un ejemplo de este fenómeno. Por eso, no se sigue que en t1 Sally no sabía que estaba pensando que el agua es a veces salada.

Otra posibilidad a mano del defensor del reemplazo conceptual es aceptar la premisa (1) y rechazar (2):

> Ahora, si resulta que el externismo de Ludlow sobre la memoria implica que para t2 Sally olvida algo que sabía en t1, entonces su objeción a la premisa (1) falla. (...) resulta difícil ver cómo el externismo de Ludlow sobre memoria muestra que de t1 a t2, "Sally no olvidó nada". El externismo sobre la memoria establece exactamente lo contrario. Establece que en t2, Sally *falla al recordar* lo que pensó en t1. Así, si en t1 Sally sabía qué estaba pensando (como defiende Ludlow), entonces en t2 Sally olvidó algo que sabía en t1.[154]

[153] Boghossian is arguably correct in asserting that [Sally does] not know at t2 what [she] knew at t1, but he is incorrect in supposing that "the only explanation" for this is that [Sally] "never knew" [her] thoughts in the first place. It is entirely consistent with the social externalist view of memory that [Sally] forgot nothing, but that the contents of [her] memories have nonetheless shifted. Indeed, this is not only *possible* according to social externalism, but given the prevalence of slow switching it should be a rather common state of affairs. (Ludlow (1995b), pp. 309-310)

[154] Now if it turns out that Ludlow's externalism about memory implies that by t2 [Sally] forgot something [she] knew at t1, then his objection to premise (1) fails. (...) it is hard to see how Ludlow's

> (*) En t, S ha olvidado que P si y sólo si (i) en t, S no recuerda que P, y (ii) para algún t' anterior a t, en t', S sabía que P. (...) Ahora, parece que la premisa (1) es una consecuencia trivial de (*). (...) Así que nuestra discusión sobre Ludlow nos deja afirmando la premisa (1) del argumento de Boghossian como teniendo estatus de obviedad. Pero las consideraciones que han emergido muestran que la premisa (2) del argumento es falsa.[155]

Creemos que la discusión entre Ludlow y Brueckner es completamente terminológica (y por lo tanto insustancial). Lo que sí es interesante es que parece que hay al menos cierta tensión entre las premisas (1) y (2) del argumento de la memoria. Ejemplos como el siguiente (el cual nada tiene que ver con semánticas externistas y transiciones lentas) así lo sugieren.

Supongamos que en t1, cuando todavía es niña, Coraline mantiene una agradable charla con Miss Forcible, una antigua actriz que vive en una casa cercana a suya. Pasan los años, y Coraline adquiere una cantidad ingente de información que la hacen creer (de hecho falsamente) que nunca habló con nadie que se llamara 'Miss Forcible': ni sus padres ni nadie en su vecindario se han topado nunca con Miss Forcible, y ninguno de ellos ha oído hablar de su existencia; todos creen erróneamente que la casa en la que vivió Miss Forcible ha estado siempre deshabitada, y achacan el recuerdo de Coraline a la imaginación infantil, le dicen que Miss Forcible no era más que una amiga imaginaria. Por eso, en t2, cuando ya es adulta, Coraline llega a la creencia de que Miss Forcible nunca existió, y achaca su recuerdo a fantasías que tuvo de niña. Pero el recuerdo es veraz: Coraline mantuvo en t1 una agradable charla con Miss Forcible.

En t1 Coraline sabe que ha hablado con Miss Forcible, no así en t2 (ya que ni siquiera lo cree). Nada falla en los procesos cognitivos que aportan información sobre su pasado a Coraline. ¿Olvida ésta que habló con Miss Forcible? Depende de cómo definamos 'olvidar'. Si estipulamos que si alguien sabe que *p* en t1 y no sabe que *p* en t2 entonces ha olvidado que *p*, se sigue que Coraline ha olvidado que habló con Miss Forcible. Pero

---

externalism about memory shows that from t1 to t2, "[Sally] forgot nothing". The externalism about memory establishes exactly the opposite. It establishes that at t2, [Sally] *fail*[*s*] *to remember* what [she] thought at t1. Thus if at t1 [Sally] knew what [she] was thinking (as Ludlow maintains), then by t2 [Sally] forgot something [she] knew at t1. (Brueckner (1997), p. 325)

[155] (*) At t, S has forgotten that P iff (i) at t, S does not remember that P, and (ii) for some t' earlier than t, at t', S knew that P. (...) Now it appears that premise (1) is a trivial consequence of (*). (...) So our discussion of Ludlow leaves us affirming premise (1) of the Boghossian argument as indeed having a platitudinous status. But the considerations that have emerged show that premise (2) of the argument is false. (Brueckner (1997), p. 326)

si esto es así, no podemos *estipular* en nuestra historia que Coraline no olvidó nada, y tampoco lo puede hacer Boghossian en el caso de Sally. Por otro lado, si creemos que alguien olvida que *p* sólo si algo *falla* en el proceso cognitivo encargado de guardar y abastecerle de información, entonces Coraline no ha olvidado que habló con Miss Forcible. Pero (1) resulta evidentemente falso, la historia de Coraline demuestra que hay otros modos de perder conocimiento más allá de olvidar; obtener evidencia que mine nuestra anterior justificación es uno de ellos, transitar de escenario podría ser otro.

Independientemente de que (3) sea verdadero o falso, (1) y (2) no pueden ser verdaderos los dos; cuál creamos que sea falso depende de cómo definamos 'olvidar'[156]. El argumento de la memoria falla a la hora de demostrar que en t1 Sally no sabía qué estaba pensando. Independientemente de que creamos que Sally sí puede recordar en t2 qué pensó en t1, el externismo semántico y el auto-conocimiento autoritativo son compatibles.

[156] También Bernecker (2004) señala este uso ambiguo de 'olvidar' en el argumento de la memoria: "No hay una única noción de olvidar que haga verdaderas las dos premisas del argumento de la memoria. (...) Comencemos con la premisa (1): si S no olvida nada, entonces lo que S sabía en t1, puede saber S en t2. Esta premisa tiene forma de una condicional. Para que el consecuente del condicional sea verdadero, el antecedente tiene que excluir no sólo *algunos* fallos de memoria, sino *todos*. (...) Por eso, el término 'olvidar' usado en el antecedente de la premisa (1) tiene que ser entendido en el sentido amplio. (...) Ahora, consideremos la premisa (2): S no olvidó nada. (...) Dada la teoría de reemplazo conceptual que acepta Boghossian, un cambio conceptual debido a transiciones lentas sí causa fallos de la memoria. Y si asumimos la noción amplia de olvidar, se sigue que las transiciones causan olvidos. Por eso el único modo que tiene Boghossian de reconciliar la teoría del reemplazo conceptual con la verdad de (2) es interpretar esta premisa como usando la noción estrecha de olvidar. (...) Resumiendo, el argumento de la memoria se apoya en un uso equívoco del término 'olvidar'." ("There is no single notion of forgetting which renders both premises of the memory argument true. (...) Let's start with premise (I): if S forgets nothing, then what S [knew] at t1, S can [know] at t2. This premise has the form of an implication. For the consequent of the implication to be true, the antecedent has to rule out not only *some* but *all* memory failures. (...) Therefore, the term 'forgetting' used in the antecedent of premise (I) has to be understood in the wide sense. (...) Next, consider premise (2): S forgot nothing. (...) Given the conceptual replacement view which Boghossian takes for granted, a conceptual shift due to unwitting slow switching *does* cause memory failure. And if we assume the wide notion of forgetting, it follows that switching causes forgetting. Therefore the only way for Boghossian to reconcile the conceptual replacement view with the truth of (2) is to read this premise as implying the narrow notion of forgetting. (...) In sum, the memory argument rests on an equivocation of the term 'forgetting'." (Bernecker (2004), p. 622))

3. DESCUBRIMIENTOS Y MEMORIA

Al menos una premisa del argumento de la memoria ha de ser, pues, falsa. Pero esta parte del trabajo no termina aquí; como hemos dicho antes, los ejemplos del tipo de Sally proponen preguntas de lo más interesantes acerca de los compromisos que podría tener el externista acerca de la memoria, y en lo que sigue nos gustaría decir algo más sobre el tema.

En este capítulo nos centraremos en una tesis defendida por John Gibbons en "Externalism and Knowledge of Content" (1996). Como hemos expuesto antes, Gibbons defiende que en t2 Sally puede saber qué pensó en t1, ya que, según él, no hay motivos suficientes para negar que posee el concepto AGUA en t2 y porque, dado que el pensamiento de segundo orden se infiere exclusivamente del pensamiento de primer orden, aquél hereda los conceptos de éste. Pero, de acuerdo con Gibbons, las cosas serían distintas si Sally adquiriera suficiente información acerca de su pasado—cambiemos un poco el ejemplo que venimos discutiendo en esta parte.

Supongamos que, después de t2, Sally sufre varias transiciones de la Tierra Gemela a la Tierra (y viceversa), y que pasa el tiempo suficiente cada vez en el entorno nuevo para adquirir (o comenzar a usar de nuevo) el concepto en cuestión. Supongamos que un día, en t3, mientras se encuentra en la Tierra Gemela, le contamos su historia. Sally descubre

que ha habitado dos entornos distintos, que hay dos sustancias distintas que ella denominaba 'agua', y que sus pensamientos y sus proferencias del término 'agua' referían a dos sustancias distintas; además, adquiere el vocabulario suficiente para poder distinguir entre agua y bi-agua cuando habla (ahora usa el término 'agua' para referirse sólo al agua y el término 'bi-agua' para referirse a la bi-agua). Pero supongamos también que no le decimos en qué entorno se encontraba en cada momento, ni tampoco en qué entorno se encuentra ahora. Así, en t3, Sally no sabe si su entorno en t1 contenía agua o bi-agua, ni tampoco sabe si su entorno en t3 contiene agua o bi-agua. Llegados a este punto, le preguntamos: 'Sally, ¿qué pensaste en t1?'

La pregunta que nos interesa aquí es: ¿puede Sally saber en t3, mediante introspección, qué pensó en t1? ¿Puede Sally saber en t3, sin ningún tipo de investigación empírica, que en t1 pensó que el agua es a veces salada? Gibbons (1996) defiende que no. Según él Sally pierde el concepto AGUA que tenía en t1 cuando descubre cómo ha sido realmente su pasado y, por eso, es incapaz de pensar aquello que pensó en t1. Porque no puede pensarlo, no puede saber que lo pensó.

3.1. EL ARGUMENTO DE GIBBONS: REEMPLAZO DE CONCEPTOS

Gibbons comienza señalando que los conceptos AGUAt3[157] y BI-AGUAt3 de Sally tienen diferentes roles funcionales, ya que en t3 Sally no acepta sus creencias sobre agua como una buena evidencia a favor o en contra de sus creencias sobre bi-agua (y viceversa)[158]. Pero, dado que en t2 Sally no distingue entre el agua y la bi-agua, según Gibbons se sigue que sus conceptos AGUAt2 y BI-AGUAt2 tienen el mismo rol funcional: en t2 Sally

[157] A lo largo de este capítulo, usaremos la expresión 'AGUAtn' como abreviatura de 'concepto sobre el agua que tiene Sally en tn'.

[158] El argumento de Gibbons también presupone que la referencia o extensión de un concepto es relevante a la hora de individuar ese concepto (esto es, el argumento presupone que el externismo es verdadero). De hecho, es porque presupone que el externismo es verdadero que también defiende que el concepto que expresa Sally en t2 cuando profiere el término 'agua' tiene un rol funcional distinto al concepto que expresa cuando en t3 profiere el término 'agua'. Si el externismo no fuera verdadero, entonces los pensamientos de Sally en t1 y t2 sobre agua y bi-agua estarían constituidos por un mismo concepto (que Sally expresaría profiriendo el término 'agua'). Y nada cambiaría si Sally descubriera que ha estado transitando de un entorno a otro. Dado que estarían constituidas por el mismo concepto, Sally podría aceptar sus creencias sobre agua como evidencia a favor o en contra de sus creencias sobre bi-agua, y cuando se le preguntara qué pensó en t1, podría responder diciendo que pensó que el agua es a veces salada.

sí acepta sus creencias sobre agua como una buena evidencia a favor o en contra de sus creencias sobre bi-agua (y viceversa). Por ello, se sigue que AGUAt2 y AGUAt3 tienen distintos roles funcionales; dado que AGUAt2 y BI-AGUAt2 tienen el mismo rol funcional, sería completamente arbitrario identificar su rol funcional con el de AGUAt3 (y no con el rol funcional de BI-AGUAt3). Como AGUAt1 y AGUAt2 son el mismo concepto, se sigue que AGUAt1 y AGUAt3 tienen distintos roles funcionales.

Gibbons se basa en esta diferencia en sus roles funcionales para defender que AGUAt1 y AGUAt3 son dos conceptos distintos. Ahora, ¿se sigue dentro de un marco externista que dos conceptos (que un mismo individuo tiene en dos momentos distintos) son dos conceptos distintos por el mero hecho de que tienen distintos roles funcionales? Es evidente que no, y así lo admite también Gibbons: "Va claramente en contra del espíritu de gran parte de la literatura externista el decir que cada vez que alguien aprende algo nuevo acerca de un individuo o una clase natural termina con un concepto diferente"[159]. Pero sí parece que dos conceptos de un mismo individuo serán distintos si tienen roles funcionales *suficientemente* diferentes (aún en el caso de que tengan la misma referencia); y esta tesis es compatible con la tesis externista principal, a saber, que algunos factores externos (factores que no tienen por qué afectar al rol funcional de los conceptos) pueden ser relevantes a la hora de individuar un concepto. De hecho, probablemente algo tendremos que decir sobre roles funcionales si queremos mantener que los conceptos TULIO y CICERÓN de un individuo son dos conceptos distintos.

Gibbons adopta explícitamente el siguiente principio sobre individuación de conceptos: Si S cree que los conceptos C y C' refieren a entidades distintas (si S cree que el enunciado 'c≠c'' expresa una proposición verdadera), entonces los conceptos C y C' de S serán dos conceptos distintos, ya que S no aceptará sus creencias en parte constituidas por C como evidencia a favor o en contra de sus creencias en parte compuestas por C'. Y

> Dado que Sally no sabe cuándo sufrió la transición, no llega a creer el bicondicional relevante 'algo es de *aquella* clase (pensando sobre la clase que llamaba 'agua' en t1) sólo si es de *esta* clase' (pensando sobre la clase que llama 'agua' ahora). Es éste exactamente el tipo de situación donde tomamos las diferencias en el rol funcional como relevantes para la individuación de conceptos

[159] It clearly goes against the spirit of much of the externalist literature to say that any time you learn something new about an individual or kind you end up with a different concept" (Gibbons (1996), p. 307).

> en el caso intrapersonal. Por eso, muchos externistas ya saben que necesitan una explicación que distinga entre AGUAt1 y AGUAt3 o BI-AGUAt3.[160]

Esto es, porque en t3 Sally no cree que su pensamiento en t1 y sus creencias actuales sobre agua son sobre la misma sustancia (porque en t3 no tiene la creencia de que algo es lo que en t1 solía llamar 'agua' si y sólo si es agua), sus conceptos AGUAt1 y AGUAt3 son dos conceptos distintos. Por eso, porque los conceptos AGUAt1 y AGUAt3 de Sally son dos conceptos distintos, Gibbons concluye que cuando en t1 Sally pensó que el agua es a veces salada, lo hizo sin emplear ninguno de los conceptos AGUA y BI-AGUA que tiene en t3; y por eso

> Si la pregunta "Sally, en t1, pensaste sobre agua o bi-agua?" es realmente una pregunta sobre cómo concibió las cosas en t1, Sally no pensó sobre el agua en ninguno de los modos empleados en la pregunta "Sally, en t1, pensaste sobre agua o bi-agua?"[161]

Y, según Gibbons, esto es suficiente para concluir que Sally no puede recordar en t3 el pensamiento que tuvo en t1. En t3 está usando conceptos que no tenía en t1 y, por lo tanto, en t3 carece de algunos conceptos que tenía en t1 y que en parte constituían el pensamiento que ahora está intentando recordar. Como en t3 carece del concepto AGUAt1, Sally no puede pensar que el agua es a veces salada *tal y como lo pensó en t1*. Y, como no puede pensar ese pensamiento, tampoco puede auto-adscribírselo; en t3 Sally no puede pensar que en t1 pensó que el agua es a veces salada en un modo tal que el concepto AGUAt1 constituya en parte ese pensamiento. Por eso, en t3, Sally no puede saber, sin ayuda de más investigación empírica, qué pensó en t1:

> Se sigue de esta solución que Sally no sabe qué pensó en t1. En cuanto en t3 emplea conceptos distintos a aquellos que emplea en t1, no puede decir o pensar lo que pensaba entonces. Pero si no puede pensarlo, no puede saberlo.[162]

---

[160] Since [Sally doesn't] know when [she] was switched, [she fails] to believe the relevant biconditional ['something is of *that* kind (thinking of the kind she called 'water' at t1) just in case it is of *this* kind' (thinking of the kind she calls 'water' now)]. This is exactly the sort of situation where we take differences in functional role as relevant to the individuation of concepts for the intrapersonal case. So, many externalists already know that they need an account that will distinguish between [WATERt1] and [WATERt3] or [TWATERt3]. (Gibbons (1996), p. 309)

[161] If [the question "Sally, at t1, did you think about water or twater?"] is really a question about how [Sally] conceived of things [at t1, Sally] did not think about water in either of the ways involved in [the question "Sally, at t1, did you think about water or twater?"]. (Gibbons (1996), p. 309)

[162] ...it follows from this solution that [Sally does] not know what [she] was thinking [at t1]. Since at t3 [she employs] concepts distinct from those [she employs] at t1, [she] cannot say or think what [she] was thinking then. But if [she] cannot think it, [she] cannot know it. (Gibbons (1996), p. 309)

3.2. INDIVIDUACIÓN DE CONCEPTOS (CRÍTICA A GIBBONS)

Creo que el siguiente esquema resume adecuadamente el argumento que propone Gibbons (1996):

(1) Los conceptos AGUAt1 y AGUAt3 de Sally tienen roles funcionales distintos, tal que Sally no sabe si es verdad que algo cae bajo la extensión de AGUAt1 si y sólo si cae bajo la extensión de AGUAt3.
(2) Por eso, los conceptos AGUAt1 y AGUAt3 de Sally son dos conceptos distintos.
(3) Por eso, en t3, Sally carece de ningún concepto que sea idéntico a AGUAt1.
(4) Sally necesita el concepto AGUAt1 para poder pensar aquello que pensó en t1 y para ser capaz de recordarlo.
(5) Por eso, en t3 Sally no puede conocer no-empíricamente que en t1 pensó que el agua es a veces salada (porque ni siquiera puede pensar y recordar ese pensamiento en el modo en el que solía hacerlo).

Estamos de acuerdo con la tesis que defiende Gibbons (que en t3 Sally no puede saber sólo mediante introspección qué pensó en t1), más adelante explicaremos por qué; pero también creemos que su argumento no es acertado. Los principios para individuación de conceptos que asume son, al menos, problemáticos y, además, el argumento que ofrece no es válido.

La premisa (3) no se sigue de la premisa (2). Supongamos que Gibbons (1996) está en lo cierto al afirmar que AGUAt1 y AGUAt3 son dos conceptos distintos, de ello no se sigue que en t3 Sally no puede repensar su pensamiento de t1. Para llegar a esa conclusión Gibbons necesita demostrar que Sally carece del concepto AGUAt1; y ofrece argumentos para apoyar que Sally adquiere dos nuevos conceptos al aprender sobre las transiciones, pero no da ninguno para apoyar que pierde sus conceptos anteriores. ¿Por qué deberíamos creer que es éste un ejemplo de reemplazo de conceptos en vez de uno de cohabitación? La falta de tal explicación resulta incluso más extraña si tenemos en cuenta que Gibbons (1996) explícitamente defiende que la situación de Sally en t2 es un ejemplo de cohabitación de conceptos, no de reemplazo. A falta de una explicación acerca de esta asimetría entre t2 y t3 el argumento simplemente no es válido.

Pero creemos que no es ése el único problema de la propuesta: creemos que los criterios para individuación de conceptos que asume Gibbons son, al menos, problemáticos. Recordemos qué le ocurre a Sally según él. En t2, antes de saber acerca de sus viajes, tiene dos conceptos distintos (AGUAt2 y BI-AGUAt2); hay dos clases naturales distintas que Sally etiqueta como 'agua' y, según el externismo de clases naturales, eso es suficiente para concluir que los dos conceptos correspondientes son distintos[163]. Ahora, de acuerdo con Gibbons, Sally adquiere dos nuevos conceptos (AGUAt3 y BI-AGUAt3) cuando descubre su historial de transiciones. No podría responder a la pregunta de si la sustancia que etiquetaba como 'agua' en t1 es la misma sustancia que en t3 conoce como 'agua', ni a la pregunta de si la sustancia que etiquetaba como 'agua' en t2 es la misma sustancia que en t3 conoce como 'bi-agua'; y, de acuerdo con Gibbons, eso es suficiente para concluir que los conceptos AGUAt1 y AGUAt3 de Sally por un lado, y sus conceptos BI-AGUAt2 y BI-AGUAt3 por el otro, son distintos. Creemos que esta conclusión es problemática; nos gustaría abogar por una teoría "más externista"[164] sobre individuación de conceptos, una tal que concluiría que Sally no adquiere ningún concepto nuevo cuando aprende acerca de sus viajes.

Supongamos que en t0, antes de ninguna transición y antes de pensar en t1 que el agua es a veces salada, Sally pensó que le gusta arrojar globos llenos de agua a la gente. Es evidente que el concepto AGUA que en parte constituye ese pensamiento en t0 es el mismo concepto AGUA que en parte constituye su pensamiento en t1. Pero supongamos también que en t3, cuando le contamos a Sally su historia, le decimos que su entorno en t0 contenía agua, y no bi-agua. ¿Son los conceptos AGUAt0 y AGUAt3 el mismo concepto? Ya hemos explicado que, según Gibbons, AGUAt1 y AGUAt3 son dos conceptos distintos; AGUAt0 y AGUAt1 son a todas luces el mismo concepto y, por eso, Gibbons está comprometido a defender que AGUAt0 y AGUAt3 son dos conceptos distintos. Bien, creemos que esta última tesis es, en el mejor de los casos, problemática.

Gibbons no ofrece ninguna condición suficiente de identidad entre conceptos, pero sí propone una necesaria: para que dos conceptos C y C' de un individuo sean el mismo concepto, éste tiene que creer que algo cae dentro de la extensión de C si y sólo si cae

[163] Por ahora asumiremos que Sally no pierde el concepto AGUA cuando adquiere el concepto BI-AGUA, y que el que adquiere no es un "concepto amalgama".
[164] Fodor (1998) es un buen ejemplo del tipo de teoría "más externista" por la cual nos gustaría abogar.

dentro de la extensión de C'. Bien, AGUAt0 y AGUAt3 tienen la misma extensión, y Sally lo sabe; Sally sí cree que algo cae dentro de la extensión de AGUAt0 si y sólo si cae dentro de la extensión de AGUAt3. Por eso, AGUAt0 y AGUAt3 guardan algo más que una "similitud general en sus roles funcionales": Sally sabe que sus creencias en parte constituidas por el concepto AGUAt0 son una buena evidencia a favor o en contra de sus creencias en parte constituidas por el concepto AGUAt3 (y viceversa); por ejemplo, en t3 Sally puede deducir de la creencia que tenía en t0 de que el agua hierve a cien grados que el agua hierve a cien grados.

La situación de Gibbons es más problemática de lo que parece. Una vez aceptamos que AGUAt0 y AGUAt3 son dos conceptos distintos, cabe preguntar si Sally sabe en t3 qué pensó en t0. Parece que la respuesta evidente es que sí, que sabe que pensó que le gusta arrojar globos llenos de agua. Pero, por otro lado, si Sally carece del concepto AGUAt0 (cosa que Gibbons ha de defender), parece que no puede pensar ese pensamiento que se está auto-adscribiendo veraz y justificadamente. Esto es raro. ¿Qué pensamiento se auto-adscribe Sally si profiere el enunciado 'en t0 pensé que me gusta arrojar globos de agua'? Si la proferencia es verdadera (y la proposición expresada está en parte constituida por el concepto AGUAt0) Sally no puede entender aquello que acaba de decir. Si Sally entiende lo que acaba de decir (y la proposición expresada está en parte constituida por el concepto AGUAt3) entonces Sally no sabe qué pensó en t0, porque la auto-adscripción que está haciendo es falsa. No creemos que ninguna de las dos opciones sea aceptable, es preferible mantener que los conceptos AGUAt0 y AGUAt3 de Sally son el mismo concepto. Pero si esto es así, entonces también los conceptos AGUAt1 y AGUAt3 de Sally son el mismo concepto. Defender que en t3 Sally no puede descubrir o aprender que en t0 pensó que le gustaba arrojar globos llenos de agua (por ejemplo, descubriendo que t0 es un escenario de agua) es bizarro; por lo tanto, en t3 Sally sí tiene el concepto AGUAt1.

3.3. ¿UNA EXPLICACIÓN "EN TÉRMINOS EPISTÉMICOS"?

También a nosotros nos parece que en t3 Sally no puede recordar que en t1 pensó que el agua es a veces salada, pero la explicación dada por Gibbons no es aceptable. Es muy

extraño asumir que en t3 Sally no tiene el concepto AGUA que tenía en t1; sería conveniente que la explicación de por qué Sally no puede en t3 recordar qué pensó en t1 no se basara en posibles incapacidades conceptuales de ésta. Así, Brueckner (1997) propone que Sally no podría recordar en t3 porque carece de la justificación necesaria:

> Éste podría ser un caso en el que S recuerda que P pero no sabe que P sobre la base de esa memoria. En este caso, la información sobre las transiciones elimina la justificación de Sally para creer que pensó en t1 que el agua es a veces salada. Parece plausible mantener que, en general, S sabe que P sobre la base de recordar que P sólo si S no tiene ninguna razón para dudar de la exactitud de su aparente memoria de que P.[165]

Hay una pequeña diferencia entre la propuesta que sugiere Brueckner (1997) y la que defenderemos nosotros, que concierne a si Sally *recuerda* en t3 que en t1 pensó que el agua es a veces salada—a diferencia de Brueckner, nosotros creemos que no. Porque 'S recuerda que p' implica 'S sabe que p'; por eso, dado que (como Brueckner) negamos que en t3 Sally *sabe* que en t1 pensó que el agua es a veces salada, también negamos que en t3 *recuerda* que en t1 pensó que el agua es a veces salada. Otra cosa es que en t3 recuerde un estado de cosas en el que pensó que el agua es a veces salada, lo cual creemos que es verdadero, ya que en t3 Sally obtiene mediante su memoria una representación de sus experiencias en t1, donde pensaba que el agua es a veces salada; pero, trayendo a colación una distinción introducida por Williamson (1995), una cosa es recordar un estado de cosas en el que *p*, y otra distinta recordar que *p*, y de que S recuerde un estado de cosas en el que *p* no se sigue que S recuerda que *p*. Sea como sea, la discusión acerca de si de 'S recuerda que *p*' se sigue 'S sabe que *p*' nos parece en gran parte terminológica.

En el siguiente capítulo esbozaremos las bases para una teoría de la memoria que nos permita mantener que en t2 Sally tiene evidencia suficiente para recordar que en t1 pensó que el agua es a veces salada, pero que no así en t3.

[165] This may be a case in which S remembers that P but does not know that P on the basis of that memory. In the current case, the information about switching defeats [Sally's] justification for believing that [she] thought at t1 that [water is sometimes salty]. It seems plausible to hold that in general, S knows that P on the basis of remembering that P only if S has no reason to doubt the accuracy of his apparent memory that P. (Brueckner (1997), p. 328)

4. BREVE ESBOZO DE UNA (PROTO)TEORÍA DE LA MEMORIA

En este capitulo esbozaremos una teoría de la memoria. No entraremos demasiado en detalles; nos contentaremos con dibujar las líneas generales de una supuesta propuesta que nos posibilite sostener cierta posición en las cuestiones sobre memoria, conocimiento, externismo y transiciones que hemos introducido en las páginas anteriores. En concreto, queremos que nuestro modelo de memoria nos capacite para mantener al menos dos cosas:

- Que Sally puede conocer en t2 sus vivencias en t1. Queremos, pues, una teoría de la memoria que se base en la cohabitación conceptual—no entraremos a justificar en este capítulo por qué abogamos por un modelo así, dejamos esta tarea para el capítulo siguiente.
- Que Sally no puede saber en t3 qué pensó en t1, y que esto es así porque ha perdido la justificación que tenía en t2.

Con este fin, mencionaremos qué debería afirmar una teoría de la memoria sobre qué determina el contenido de un recuerdo y qué justifica una creencia basada en la memoria. La propuesta que esbozaremos distingue entre memoria semántica y memoria

episódica—esta distinción no nos es interesante sólo por meras cuestiones de taxonomía mnemónica; defenderemos que estas "dos memorias" tienen una naturaleza epistémica distinta.

Lo que define la memoria episódica y la memoria semántica, lo que las diferencia la una de la otra, es su distinta función, su objetivo. El objetivo de la memoria episódica es proporcionarle al sujeto información y conocimiento acerca de su pasado, acerca de los distintos eventos autobiográficos que ha vivido. Por contra, el objetivo de la memoria semántica es proporcionarle al sujeto conocimiento general que ya había adquirido con anterioridad. Son ejemplos de memoria episódica el recuerdo de nuestra primera bicicleta, cuando Idoia recuerda que anoche terminó de leer *Orgullo y Prejuicio* o cuando Manolo recuerda que esta mañana se ha caído de la bicicleta al tropezar con una paloma. Por contra, la memoria semántica nos permite recordar que Cervantes escribió el Quijote, la lista de los Reyes Godos, o que Yakarta es la capital de Indonesia.

A primera vista, parece que la memoria episódica y la memoria semántica también se diferencian en al menos otras dos cosas. Primero, y ya diremos algo más sobre esto, la memoria episódica puede ser tanto proposicional como no-proposicional, mientras que la memoria semántica siempre es proposicional. Segundo, cuando "S recuerda que *p*" es un ejemplo de memoria episódica, necesariamente S ha de recordar en qué contexto aprendió que *p*, no así cuando "S recuerda que *p*" es un ejemplo de memoria semántica. Si Idoia recuerda que anoche terminó de leer *Orgullo y Prejuicio*, entonces necesariamente recuerda el contexto en el cual aprendió que anoche terminó de leer *Orgullo y Prejuicio* (lo aprendió anoche, al acabar de leer *Orgullo y Prejuicio*), pero en cambio Idoia puede recordar que Jane Austen es la autora de *Orgullo y Prejuicio* aun y cuando no recuerda en qué contexto lo aprendió.

Algunos psicólogos hacen un uso ambiguo de estos dos términos 'memoria episódica' y 'memoria semántica', refiriéndose algunas veces a dos procesos cognitivos. No lo haremos así nosotros; de hecho, intentaremos ceñirnos a cuestiones de semántica, capacidad conceptual y epistemología, evitando en lo posible entrar en cuestiones de psicología. En lo que queda de capítulo intentaremos explicar en líneas generales qué capacidades conceptuales y epistémicas necesita un sujeto S para recordar que *p*; brevemente, defenderemos que estos dos tipos de memoria muestran diferencias en su

naturaleza epistémica, porque cuando uno recuerda episódicamente que *p* hay un elemento de carácter mnemónico en su justificación, pero no así cuando uno recuerda semánticamente que *p*.

4.1. MEMORIA EPISÓDICA

Hay un elemento que juzgamos central en la epistemología de la memoria episódica: la representación mnemónica. En pocas palabras, la memoria episódica provee al sujeto una representación, y es sobre la base de esa representación que el sujeto puede llegar a recordar que *p*. Pero queremos dejar clara una cosa: no defendemos que la representación mnemónica sea el objeto de la memoria, lo que defendemos es que es la evidencia en la base del recuerdo episódico que tiene uno de que *p*. Por eso, porque creemos que no recordamos representaciones, sino objetos, hechos o verdades, nuestra propuesta no se enmarca dentro de un modelo representacional de la memoria[166]. Y porque defendemos que la representación mnemónica constituye la evidencia que tiene uno para recordar episódicamente que *p*, también defendemos que cuando uno sabe que *p* sobre la base de su memoria episódica, hay en su justificación un elemento que tiene naturaleza mnemónica: la representación mnemónica.

Diferenciemos aquí entre recordar una situación, un evento, un estado de cosas o un objeto, por un lado, y recordar que *p*, recordar que esto o lo otro sucedió; a esto último le llamaremos 'memoria proposicional', a lo primero, 'memoria no-proposicional'. Un sujeto S recuerda un estado de cosas (un hecho o un objeto) si y sólo si tiene una representación mnemónica de ese estado de cosas. Así, en t2 S recuerda un estado de cosas (un objeto o un hecho) W que tuvo lugar en t1 (tiene el recuerdo de un estado de cosas W) si y sólo si:

- En t2 S tiene una representación de W.
- En t1 S percibe el estado de cosas W.

[166] Para un ejemplo paradigmático de teoría representacional de la memoria, véase Martin y Deutscher (1966); para una breve caracterización, Audi (1997) o Bernecker (2008)

- Hay una relación causal *directa* entre la experiencia de W que tiene S en t1 y la representación de W que tiene S en t2.

No queremos entrar en demasiados detalles, pero sí nos gustaría por lo menos decir algo acerca de la naturaleza de esa *representación mnemónica* que venimos mencionando. La representación mnemónica es una imagen cualitativa que viene asociada al recuerdo cuando alguien recuerda algo episódicamente. La representación mnemónica que tiene uno cuando recuerda episódicamente W es, digámoslo así, como la experiencia perceptiva que tendría si percibiera W. Así, en cierto sentido y salvando las distancias, recordar nuestra primera bicicleta es como percibir esa bicicleta[167]. De hecho, dada la segunda condición que hemos expuesto (que alguien recuerda W sólo si percibió W), se sigue que la memoria episódica tiene su origen en una percepción: uno puede recordar episódicamente un estado de cosas sólo si antes ha percibido ese estado de cosas[168].

Ahora, la representación mnemónica no *refiere* a esa percepción original, ya hemos mencionado antes que no es la nuestra una teoría representacional de la memoria. Quisiéramos mencionar también otra cuestión. La representación mnemónica no es una "copia" o un "darse de nuevo" de la percepción original. Uno puede tener una representación mnemónica de su primera bicicleta sin estar por ello en posición de saber y recordar cuántos radios tenia la rueda delantera de esa bicicleta, a pesar de que, es de suponer, su representación mnemónica representará la bicicleta como teniendo radios en la rueda delantera. Por contra, si uno percibe en t su primera bicicleta, está en t en posición de saber cuántos radios tiene en la rueda delantera, no tiene más que contarlas.

La representación mnemónica se parece más a un esquema condicionado conceptualmente que a una copia de una percepción que se tuvo antaño—y es la relación causal con el estado de cosas y su percepción original la que "carga conceptualmente" la representación mnemónica. Uno puede tener una representación mnemónica R sobre un objeto o, tal que la representación mnemónica "se parece" más al objeto o'. No se sigue que R referirá a o'; el que haya una cadena causal entre o y R

[167] Evidentemente, la percepción de un objeto y el recuerdo de un objeto son fenoménicamente diferentes; ya Locke, Hume o Reid intentaron determinar en qué se diferencia la percepción de un objeto de su recuerdo. Pero precisar en qué consiste esta diferencia fenoménica resulta extremadamente difícil, y no es nuestra intención aportar nada a esta discusión. Por eso, no entraremos en estas cuestiones.

[168] Y lo mismo serviría para la memoria episódica proposicional que caracterizaremos más abajo: alguien puede recordar episódicamente que *p* sólo si ha percibido antes un estado de cosas en el que se da que *p*.

asegura que haya una relación de referencia entre ellos, da igual que R sea cualitativamente más parecida a o'. Manolo por ejemplo recuerda la paloma que esta mañana lo tiró de la bicicleta, porque guarda una representación mnemónica de esa paloma. Y la representación es sobre *esa* paloma, y no sobre otra paloma cualquiera, porque hay una relación causal entre la percepción original de la paloma y la representación. Es posible que la representación que guarda Manolo se parezca más a la paloma que lo tirará de la bicicleta mañana, pero da igual, su recuerdo es el recuerdo de una paloma que ya sufrió.

Además, hemos dicho que la relación entre el estado de cosas y la representación mnemónica que causa ha de ser *directamente* causal. Lo que queríamos decir con eso es que hay una relación causal entre la percepción original de W en t1 y la representación de W que guarda S en t2, y que la representación de W que tiene S en t2 será un recuerdo de W sólo si en el transcurso del tiempo que va de t1 (cuando S percibe W) a t2 (cuando tiene el recuerdo de W) S en ningún momento ha perdido la representación de W que tenía. Supongamos, por ejemplo, que en un tiempo intermedio entre t1 y t2 S le cuenta a X algo relativo a W y que, más tarde, S olvida por completo todo lo que tiene que ver con W. Si en t2 X le cuenta a S aquello relativo a W que éste le contó antes, entonces en t2 S tiene una representación de W que está causalmente relacionada con la percepción de W que tuvo en t1. Pero no queremos decir que en t2 S tiene un recuerdo de W. Defendemos que en este caso en t2 S no tiene un recuerdo de W porque la relación causal con la experiencia que tuvo en t1 no es *directa*, porque en un momento entre t1 y t2 S perdió la representación de W que tenía[169].

Pasemos ahora a la memoria proposicional episódica, cuando uno recuerda episódicamente que *p*. En t2 S recuerda episódicamente que *p* si y sólo si:

- En t2 S tiene una representación mnemónica del estado de cosas W en el cual se daba que *p*.
- S cree que *p* sobre la base de esa representación mnemónica.[170]

[169] La mayoría de autores acude a lo que llaman 'huellas', 'rastros' o 'trazos' (en inglés 'traces') para explicar lo que nosotros llamamos 'relación causal *directa*'. No nos interesan aquí estos detalles, no entraremos a discutir sobre la existencia o no de estos *trazos*, o sobre cuál es su naturaleza.

[170] Llamemos aquí la atención sobre un detalle que suena a perogrullo: la memoria episódica es episódica. Esto es, proporciona al sujeto información sobre episodios concretos de su autobiografía. No se sigue, por

Esto es, defendemos un modelo evidencial de la memoria episódica. Siguiendo a Williamson (1995), creemos que una cosa es recordar un estado de cosas en el que *p* y otra distinta recordar que *p*. De que S recuerde un estado de cosas en el que *p* no se sigue que S recuerde que *p*; S podría no tener alguno de los conceptos que constituye la proposición de que *p*, o S podría erróneamente creer que su recuerdo no es tal, que es un "recuerdo falso", una imaginación de un estado de cosas que nunca se dio. Además, nosotros proponemos que "el recuerdo de un estado de cosas en el que *p*" es la evidencia que tiene uno para poder recordar que *p*.

4.2. MEMORIA SEMÁNTICA

Como hemos dicho, a diferencia de la memoria episódica, la memoria semántica no nos proporciona información acerca de nuestras vivencias pasadas, su función es posibilitar que recuperemos las creencias que hemos adquirido antes. Así, por ejemplo, cuando uno recuerda que Cervantes escribió el Quijote o que los aguacates son un tipo de hortaliza, lo está recordando semánticamente. Más arriba hemos mencionado dos características que diferencian la memoria semántica de la episódica. Primero, a diferencia de la memoria episódica, uno no tiene por qué recordar en qué contexto aprendió que *p* para poder recordar semánticamente que *p*; uno no tiene por qué recordar que descubrió que Cervantes escribió el Quijote leyéndolo en la tapa de un libro para poder recordar que Cervantes escribió el Quijote. Además, la memoria semántica es siempre proposicional. Uno puede tener un recuerdo (episódico) de su primera bicicleta o de la paloma que atropelló esta mañana, pero no un recuerdo (ni semántico ni episódico) de Cervantes o Yakarta (no al menos si no es contemporáneo de Cervantes o si no ha estado nunca en Yakarta: en estos casos tendría un recuerdo episódico de Cervantes o Yakarta). Uno sólo puede recordar los objetos que ha percibido y, por eso, dado que hay una relación

lo tanto, que toda información de carácter autobiográfico que recuerde un sujeto sea producto de su memoria episódica. Cuando uno por ejemplo recuerda la fecha de su nacimiento no está recordando su nacimiento episódicamente, es la memoria semántica la encargada de proporcionar tal información. O supongamos por contra que alguien recuerda que ha comido trufa sin poder recordar ningún episodio concreto de haber comido trufa. También en estos casos es la memoria semántica la que le proporciona ese recuerdo. El sujeto tiene al principio recuerdos episódicos de haber comido trufa; es posible que, con el tiempo, el sujeto guarde esta información de forma semántica y que olvide los episodios concretos de haber comido trufa. En estos casos, recordará semánticamente haber comido trufa.

causal entre la percepción del objeto y el recuerdo posterior de ese objeto, uno puede recordar un objeto sólo episódicamente, y no semánticamente (en los siguientes párrafos entraremos a decir un poco más sobre memoria semántica, representaciones y percepción).

A diferencia de la memoria episódica, la memoria semántica no se basa en representaciones mnemónicas, los recuerdos semánticos de uno no están necesariamente asociados con representaciones mnemónicas de este tipo. Cuando uno recuerda que Cervantes escribió el Quijote, por ejemplo, no tiene necesariamente asociado al recuerdo una imagen perceptiva de Cervantes escribiendo el Quijote. Y es que, al no ser recuerdos de nuestras vivencias pasadas, nuestros recuerdos semánticos no tienen necesariamente en su origen una percepción de aquello que recordamos. Por lo tanto, no hay necesariamente una relación causal entre nuestro recuerdo semántico y una percepción original; lo que sí hay es una relación causal entre nuestro recuerdo semántico y una creencia original. Si alguien recuerda semánticamente que *p*, entonces, necesariamente, ya antes creía que *p*. La memoria semántica no genera conocimiento o creencias, pone en manos del sujeto el conocimiento y las creencias que éste ya tenía.

Dado que la función de la memoria semántica es simplemente "traernos a la mente" las creencias que teníamos antes, este tipo de memoria no aporta nada a la justificación que pueda tener el sujeto para sostener esa creencia; cuando uno sabe que *p* sobre la base de su memoria semántica, no hay en su justificación ningún elemento de naturaleza mnemónica. Así, como dice Burge (1993a) sobre el papel que juega en las inferencias deductivas lo que él llama 'memoria preservativa',

> Basarse en la memoria [semántica] ni siquiera añade a la fuerza justificatoria de la justificación deductiva. (...) Su rol en la justificación se deriva de lo que preserva.[171]

Esto es, sea en el marco de una inferencia deductiva o de un recuerdo de algo aprendido antaño, la memoria semántica se limita a preservar la justificación de la creencia que aporta al sujeto—la memoria semántica no proporciona justificación, sólo la preserva. Así, el sujeto hereda la justificación que tenía al principio aun y cuando no recuerda cuál era su evidencia original.

[171] Reliance on memory does not even add to the justificational force of the deductive justification. (...) Its role in justification derives from what it preserves. (Burge (1993a), p. 463)

Supongamos, por ejemplo, que en la escuela Idoia aprende que Jane Austen escribió *Orgullo y Prejuicio*. Tiene evidencia suficiente, porque así se lo dice su profesora, así lo lee en su libro de texto sobre literatura, y además ve varias copias de *Orgullo y Prejuicio* (en las que pone que el autor es Jane Austen). Ahora, años más tarde, Idoia recuerda semánticamente que Austen escribió *Orgullo y Prejuicio*. Pero no recuerda en qué contexto adquirió esta creencia, ni siquiera cuál era la evidencia original que tenía a favor de esta creencia. Supongamos, además, que a lo largo de su vida Idoia no ha adquirido más evidencia que justifique su creencia de que Austen escribió *Orgullo y Prejuicio*, o que en el momento que recuerda este hecho es incapaz de recordar evidencia alguna que haya adquirido a su favor. ¿Sabe Idoia que Austen escribió *Orgullo y Prejuicio*? ¿Está justificada su creencia? Sí, porque la memoria semántica preserva la justificación de su creencia original; Idoia no tiene por qué recordar cuál era la evidencia que tenía originariamente para que ahora su recuerdo esté suficientemente justificado.

4.3. DEFENSA DE LA ASIMETRÍA

Muy brevemente, acabamos de esbozar una (proto)teoría de la memoria, según la cual en los casos en los que uno recuerda episódicamente que *p* sí hay un elemento justificatorio de naturaleza mnemónica en la base de su creencia, pero no así cuando uno recuerda semánticamente que *p*. Alguien podría protestar que esta asimetría no está justificada; en esta subsección diremos algo (poco) en favor de esta asimetría. En varias ocasiones apelaremos a nuestras intuiciones sobre si alguien está justificado o no en determinadas circunstancias; somos conscientes de que algunas de esas intuiciones distan de ser claras pero, hasta donde llegamos a ver, en muchos casos son la mejor guía que tenemos para decidir en qué situaciones debemos postular elementos mnemónicos justificatorios y en qué situaciones no.

Defendamos primero que cuando uno recuerda episódicamente que *p*, sí hay algún elemento de naturaleza mnemónica en su justificación para creer que *p* (según la teoría que hemos esbozado, una representación mnemónica). Primero, creemos que este hecho

se sigue de qué justificaciones aceptamos normalmente como buenas. Supongamos que Joana dice que ayer corrió varios kilómetros. Ainara le pregunta: '¿Cómo lo sabes?'. Joana responde: 'Mi chándal está sudado, Gorka me acaba de enseñar algunas fotos en las que aparezco corriendo, *y además lo recuerdo*'. Aceptamos el 'y además lo recuerdo' como justificación, cuando a alguien se le pide que justifique una creencia sobre algún episodio que ha vivido y responde que lo recuerda, aceptamos su explicación por válida. Es más, parece evidente que si la única evidencia que tuviera Joana fuera que su chándal está sudado y que tiene fotos en las que aparece corriendo, que si Joana no recordara además que corrió, entonces la justificación total de su creencia sería menor que la que tiene. Solemos aceptar los recuerdos episódicos como justificaciones válidas, y esto es un buen indicio para pensar que cuando uno recuerda episódicamente que *p*, hay algún elemento de naturaleza mnemónica en su justificación.

Hay, además, un ejemplo que apela a nuestras intuiciones que, creemos, proporciona indicios en esta dirección. Supongamos que Joana no ha corrido ni siquiera cien metros en toda su vida. Un día, un científico loco la secuestra y manipula su cerebro, para implantarle una imagen mnemónica falsa, un "falso recuerdo". Joana despierta y tiene la imagen mnemónica de ella misma corriendo ayer, cree recordar que, ayer, justo después de despertarse, se puso el chándal y corrió varios kilómetros. Tenemos la fuerte intuición de que Joana tiene justificación para creer que ayer por la mañana corrió varios kilómetros; esto sólo se explica porque la "falsa memoria" de Joana (la imagen mnemónica que le han implantado) tiene cierta fuerza justificatoria.

Es más, opinamos que la memoria episódica no preserva la justificación, que propone ella misma un elemento justificatorio nuevo. Porque, si esto no fuera así, no podríamos explicar cómo es que Joana tiene justificación en el ejemplo en el que un científico loco le ha implantado memorias falsas. La justificación que tiene Joana no puede ser cierta "justificación original" que se ha preservado, porque no hay nada que justifique su creencia que sea anterior a la falsa memoria implantada.

Pasemos a la memoria semántica. Expliquemos primero por qué creemos necesario mantener que la memoria semántica es preservativa. Es común que uno recuerde semánticamente que *p*, pero que no sea capaz de proporcionar evidencia a favor de *p*, entre otras cosas, porque no recuerda (episódicamente) en qué contexto aprendió que *p*.

Hay varios ejemplos cotidianos de este fenómeno. Es común que uno recuerde que la capital de Indonesia es Yakarta, que el agua se hiela a cero grados o que Napoleón perdió la batalla de Waterloo, sin que sea capaz de proporcionar evidencia que justifique esas creencias. Uno asume, por ejemplo, que aprendió estas cosas en el colegio, y que tuvo buena justificación para adoptar esas creencias. La cuestión es que en esos casos *decimos* que uno *sabe* que la capital de Indonesia es Yakarta o que Napoleón perdió en Waterloo, y esto no podría ser así si no tuviera justificación. La interpretación más simple viene a ser que la memoria ha preservado esa justificación que tuvo uno, a pesar de que no tiene acceso a la justificación que tiene. La justificación no es transparente (o al menos no tanto).

Por supuesto, una alternativa es responder que es la memoria semántica la que ha aportado justificación, que la memoria semántica es como la episódica; uno tiene justificación para creer que *p* si tiene el recuerdo semántico de que *p*. Pero esto no puede ser así, si no, las creencias se auto-justificarían, bastaría con que uno creyera que *p* para que tuviera justificación de algún tipo para creer que *p*. Esto es absurdo, hay creencias que no están justificadas. De hecho, es por eso que cuando alguien dice que Yakarta es la capital de Indonesia y se le pregunta por la evidencia que tiene para creer tal cosa, no aceptamos como buena su respuesta de 'Lo recuerdo'.

Por todo ello, hay motivos para suponer que la memoria episódica y la memoria semántica difieren en su naturaleza justificatoria.

5. PREDICCIONES DE LA (PROTO)TEORÍA

En el capítulo anterior hemos esbozado una (proto)teoría de la memoria; hemos diferenciado entre memoria episódica y memoria semántica, las hemos descrito de forma muy breve, señalando cuáles son los rasgos generales que las caracterizan. Entre otras cosas, hemos defendido que cuando uno recuerda episódicamente que *p* hay un elemento de carácter mnemónico que justifica su creencia (la representación mnemónica), pero que no sucede así cuando uno recuerda semánticamente que *p*. A lo largo de este capítulo expondremos cuáles son las predicciones de nuestra (proto)teoría para los ejemplos de transición lenta que hemos presentado anteriormente.

5.1. MEMORIA EPISÓDICA Y TRANSICIONES LENTAS

Más adelante nos centraremos en ejemplos de memoria semántica, en esta sección nos limitaremos a describir qué puede recordar Sally episódicamente y qué no tanto en t2 como en t3. Con el marco de memoria episódica que hemos esbozado más arriba en mente, es así cómo entendemos nosotros el ejemplo de Sally:

(a) Las transiciones lentas son ejemplos de cohabitación conceptual y, por eso, en ningún momento pierde Sally su concepto AGUA (tampoco pierde su concepto BI-AGUA cuando descubre que ha estado transitando). Además, Sally no adquiere ningún concepto al aprender sobre su historial de viajes. En t3 tiene los mismos conceptos AGUA y BI-AGUA que tiene en t2.

(b) Tanto en t2 como en t3, Sally recuerda su situación en t1, en la cual pensó que el agua es a veces salada; Sally tiene una representación mnemónica del evento que acontece en t1. Este recuerdo, esta representación, es parte de la evidencia que tiene para llegar a conocer que en t1 pensó que el agua es a veces salada.

(c) El recuerdo que tiene Sally de t1 es en t2 evidencia suficiente para saber que en t1 pensó que el agua es a veces salada. No lo es en t3.

Ya hemos explicado en el tercer capítulo por qué creemos que (a) es verdadero. Una vez aceptamos que las transiciones lentas son casos de cohabitación conceptual, entendemos que es muy poco plausible defender que Sally adquiere o pierde ningún concepto cuando aprende sobre su historial de viajes. Por eso, dado que creemos que tanto en t2 como en t3 tiene los mismos conceptos AGUA y BI-AGUA, defendemos que tanto en t2 como en t3 puede Sally aprehender proposiciones en parte constituidas por cualquiera de estos dos conceptos. Concretamente, tanto en t2 como en t3 puede pensar que el agua es a veces salada, o que en t1 pensó que el agua es a veces salada (en la siguiente sección diremos más sobre cómo es posible que en la Tierra Gemela Sally tenga pensamientos sobre agua, o qué determina en la Tierra Gemela que un pensamiento sea un pensamiento de agua o de bi-agua).

Una vez hemos aceptado que tanto en t2 como en t3 Sally tiene los conceptos AGUA y BI-AGUA, (b) no es más que una consecuencia de la teoría de memoria episódica que hemos esbozado en el capítulo anterior. En t1 Sally piensa que el agua es a veces salada, esa experiencia causa una representación que queda guardada en su memoria. Si la memoria de Sally funciona correctamente, si Sally no olvida qué sucedió en t1, tanto en t2 como en t3 puede "traer a la mente" esa representación sobre t1 que se creó en t1. Por lo tanto, tanto en t2 como en t3 Sally tiene un recuerdo de su situación en t1 y, como

hemos dicho, ese recuerdo, esa representación mnemónica, es (parte de) la evidencia que tiene en t2 y t3 para saber y recordar que en t1 pensó que el agua es a veces salada.

En lo que queda de sección intentaremos argumentar a favor de (c). Creemos que en t2 Sally sí puede saber qué pensó en t1, y que no así en t3, porque en t2 y t3 se encuentra en situaciones epistémicas diferentes. En lo que sigue intentaremos defender esta asimetría entre t2 y t3.

Primero, creemos que con el marco sobre memoria episódica que hemos esbozado antes podemos dar una explicación de por qué en t2 Sally sí puede saber qué pensó en t1. Como hemos defendido, en t2 tiene un recuerdo de su situación en t1, y ese recuerdo es (parte de) la evidencia que tiene para poder saber que en t1 pensó que el agua es a veces salada. Parece más que plausible (si se quiere, estipulable) que Sally infiere únicamente de ese recuerdo un pensamiento que expresaría profiriendo el enunciado 'en t1 pensé que el agua es a veces salada'. No creemos que haya ningún argumento suficiente para creer que en t2 Sally carece del concepto AGUA y, dado que infiere este pensamiento *exclusivamente* de aquella representación que tiene a mano, creemos que el pensamiento que infiere hereda los conceptos que son parte de la representación de t1 que guarda en su memoria, entre otros, el concepto AGUA. En t2 Sally piensa que en t1 pensó que el agua es a veces salada y lo hace sobre la base de su memoria. Por eso, en t2 Sally recuerda y sabe que en t1 pensó que el agua es a veces salada.

Pero las cosas cambian en t3. En t3 Sally aprende que ha estado rodeada de dos sustancias distintas y, por ello, también aprende que ha estado pensando y hablando acerca de dos sustancias distintas. Esta información que adquiere tiene un impacto epistémico en Sally, tal que ya no puede saber que en t1 pensó que el agua es a veces salada sólo sobre la base de la representación de t1 que guarda. ¿Por qué? Bien, creemos que distintos ejemplos[172] sugieren que el siguiente principio (RA'') es verdadero:

[172] Supongamos que antes de t3, antes de que descubra que ha estado transitando entre la Tierra y la Tierra Gemela, mientras se encuentra en la Tierra Gemela, le enseñamos a Sally un vaso lleno de bi-agua. Le preguntamos '¿Qué es esto?', a lo que responde 'Agua'. Dado que está hablando castellano gemelo, Sally ha respondido correctamente que eso es bi-agua. Parece evidente que sabe que el vaso contiene bi-agua, y que sabe esto porque ve que el vaso está lleno de un líquido transparente, inodoro e insípido. Ahora, en t3, una vez que le hemos contado su historia, le enseñamos a Sally el mismo vaso lleno de bi-agua. Sally sabe que es el mismo vaso que vio antes, pero desconoce si se encuentra en la Tierra o en la Tierra Gemela. ¿Puede saber ahora Sally que el vaso está lleno de bi-agua sobre la base de que el vaso contiene un líquido transparente, inodoro e insípido? Creemos que no. Sally sabe que esa evidencia es

(RA'') Si S cree que la evidencia E es compatible con un escenario en el cual *q* es verdadero y *p* es falso, y S cree que un escenario en el que *q* es verdadero y *p* es falso es relevante, entonces S no puede conocer que *p* sólo sobre la base de la evidencia E.

(RA'') explica por qué Sally no puede saber en t3 que en t1 pensó que el agua es a veces salada sólo sobre la base de su memoria. El antecedente de (RA'') se da en t3: Sally cree que el recuerdo de t1 que guarda, su evidencia, es compatible con un escenario de bi-agua; cree que su recuerdo es compatible con que en t1 estuviera pensando que la bi-agua es a veces salada. Además, también cree que un escenario de bi-agua es relevante, es compatible con todo lo que sabe que en t1 estuviera en la Tierra Gemela pensando acerca de la bi-agua. Por eso, Sally no puede saber que en t1 pensó que el agua es a veces salada sólo sobre la base de su recuerdo de t1.

En resumidas cuentas, una vez interpretamos las transiciones lentas como ejemplos de cohabitación conceptual, no vemos ningún motivo para negar que en t2 Sally pueda recordar episódicamente que en t1 pensó que el agua es a veces salada, o que en t1 el

---

compatible con que el líquido sea agua. Necesita más evidencia para poder saber que el vaso contiene bi-agua. (Alguien podría tener reticencias para aceptar que Sally sabe en t2 que eso es bi-agua—concretamente, estamos pensando en filósofos como Falvey que defienden que en una transición lenta el sujeto reemplaza su concepto original por un concepto amalgama). Bueno, si es el caso, supóngase que Sally nunca ha transitado de una tierra a otra, que siempre ha vivido en la Tierra rodeada de agua, y que no sabe que el agua es H_2O. Cualquiera que no peque de escepticismos patológicos aceptará que Sally sabe que eso es agua. Pero supongamos que un día mentimos a Sally; le decimos que ha estado transitando de un escenario de agua a uno de bi-agua, lo que de hecho es falso. Y Sally nos cree. En un escenario así, Sally no puede saber, después de haber adquirido la creencia falsa de que ha estado transitando de un escenario a otro, que eso es agua. Porque ella cree que podría ser bi-agua.)

Propongamos ahora un ejemplo que nada tiene que ver con transiciones lentas, basado en el ejemplo de graneros falsos hecho famoso por Goldman que mencionamos en la primera parte. Supongamos que, en t1, Alvin da un paseo por el campo, y que se encuentra con lo que de hecho es un granero. Alvin ve a lo lejos lo que parece un granero y, sobre la base de esa experiencia visual, infiere que hay un granero allí. Supongamos que el entorno de Alvin no contiene ningún granero falso, y que se cumplen todas las condiciones necesarias para que Alvin sepa sobre la base de su percepción que hay un granero allí. En t1 Alvin sabe que hay un granero allí sobre la base de su evidencia perceptiva. Supongamos ahora que pasan los años, y que se encuentra una infinidad de veces con lo que de hecho son graneros falsos. Todas esas veces (muchas, tantas como se quiera) llega a la falsa creencia de que hay un granero allí sobre la base de su evidencia perceptiva. Supongamos ahora que le contamos a Alvin su historia, que le contamos que una infinidad de veces ha llegado a la falsa creencia de que hay un granero allí al encontrarse con lo que de hecho era un granero falso. Alvin no sabe cuándo se encontró con un granero auténtico y cuándo con uno falso; concretamente desconoce si el granero que vio en t1 es uno auténtico o uno falso. Le llevamos al lugar donde en t1 vio el granero—sabe que es el lugar donde en t1 vio lo que parecía un granero. ¿Puede Alvin saber, sobre la base de su percepción, que hay un granero allí? Creemos que no. Una vez que Alvin sabe que la percepción de algo que parece un granero le ha llevado demasiadas veces a creer falsamente que hay un granero allí, no puede basarse en esa misma evidencia para saber que hay un granero allí.

agua estaba fría, o que en t1 arrojaba globos llenos de agua por pura diversión. La cosa cambia una vez descubre que ha estado transitando de un entorno que contiene agua a un entorno que contiene bi-agua. En un escenario tal Sally no puede saber que en t1 pensó que el agua es a veces salada, o que en t1 el agua estaba fría. Pero no lo puede saber no porque haya perdido algún concepto que tenía (como defendía Gibbons (1996)), sino porque ha perdido parte de la justificación que tenía. Párrafos más arriba hemos dibujado las líneas generales de una teoría de la memoria episódica que nos explica qué justificación ha perdido Sally.

5.2. MEMORIA SEMÁNTICA Y TRANSICIONES LENTAS

Pasemos ahora a las predicciones sobre qué puede recordar semánticamente Sally después de transitar a la Tierra Gemela. Supongamos que en t2, una vez ha transitado a la Tierra Gemela y ha adquirido el concepto BI-AGUA, pero antes de descubrir que ha estado viajando de un escenario a otro, su memoria semántica le trae a Sally un pensamiento que expresaría profiriendo el siguiente enunciado:

(hielo) El agua se hiela a cero grados.

¿Cuál es el contenido de (hielo)? Nos basamos en un modelo de cohabitación conceptual; en t2, pues, Sally tiene tanto el concepto AGUA como el concepto BI-AGUA—en principio el pensamiento expresado por una proferencia de (hielo) podría estar constituido por cualquiera de esos dos conceptos. El resultado al que queremos llegar nosotros es el siguiente: hay casos paradigmáticos en los cuales Sally tiene un pensamiento en parte constituido por el concepto AGUA, casos paradigmáticos en los cuales Sally tiene un pensamiento en parte constituido por el concepto BI-AGUA, y casos en los cuales el pensamiento de Sally no estará compuesto ni por el concepto AGUA ni por el concepto BI-AGUA (defenderemos que estos casos se moverán entre la indeterminación y los conceptos amalgama).

Por lo tanto, la primera pregunta que intentamos responder en esta sección es la siguiente: cuando un sujeto tiene dos conceptos distintos C y C' entre los cuales no

distingue y que expresa profiriendo el término 't', ¿cuál es el contenido de un pensamiento que expresaría profiriendo un enunciado que contiene el término 't', cuál de los dos conceptos C y C' constituye en parte ese pensamiento? Los siguientes puntos esbozan más o menos nuestras opiniones al respecto:

- Las distintas proferencias de un enunciado que contiene el término 't' pueden expresar pensamientos distintos, algunos de ellos en parte constituidos por el concepto C, otros por el concepto C' y otros por algún tipo de concepto amalgama (también hay casos en los que el pensamiento no tiene un contenido determinado).
- Son varios los factores que determinan el contenido de un pensamiento. Entre ellos se encuentran: otros estados mentales de los que en parte se infiere el pensamiento, la intención del sujeto de ceñirse a las prácticas de una comunidad de hablantes, alguna relación de ostensión entre el sujeto y alguna clase natural u objeto en su entorno.
- Si entre los factores que determinan el contenido de un pensamiento p el concepto C (y sus instancias) tiene una presencia dominante en comparación con el concepto C' (y sus instancias), entonces p estará en parte constituido por el concepto C; si entre los factores que determinan el contenido de p el concepto C' (y sus instancias) tiene una presencia dominante en comparación con el concepto C (y sus instancias), entonces p estará en parte constituido por el concepto C'; si entre esos factores ni C (ni sus instancias) ni C' (ni sus instancias) tienen una presencia dominante, entonces el pensamiento no contiene ni el concepto C ni el concepto C'; algunas veces p estará constituido por un concepto amalgama y otras será un caso de indeterminación[173,174].

[173] Creemos que es posible decir algo sobre en qué tipo de casos p estará compuesto por un concepto amalgama y en que tipo de casos será un ejemplo de indeterminación. Resulta intuitivo decir que alguien emplea un concepto amalgama cuando expresa un pensamiento general, pero no así cuando el pensamiento expresado es particular. Así, los ejemplos de indeterminación serán más comunes cuando uno intenta referirse a un objeto o cuando el pensamiento se base en alguna ostensión, y los ejemplos de pensamientos constituidos por conceptos amalgama serán más comunes cuando uno intenta decir algo sobre un tipo de cosas.

[174] Quizás se nos critique, debido a los casos de indeterminación, que esta teoría que esbozamos es demasiado ambigua. Al fin y al cabo, venimos a decir, en los casos en los que predecimos indeterminación no está claro qué esta diciendo o pensando Sally, alguien puede pensar que no adscribir contenidos específicos a casos concretos podría ser un problema para la teoría. La cuestión es que creemos que hay casos en los que lo más adecuado es pronosticar que son ejemplos de indeterminación. Nuestras adscripciones de conceptos y pensamientos se basan en nuestras intuiciones acerca de qué diríamos en tales casos que está pensando el sujeto. Creemos que hay casos paradigmáticos en los que no

Esto es, creemos que distintas proferencias de (hielo) pueden expresar pensamientos en parte constituidos por el concepto AGUA, por el concepto BI-AGUA, o por algún tipo de concepto amalgama, dependiendo de cuáles sean los factores relevantes que determinan el contenido de cada pensamiento en cuestión.

Con esto en mente, resulta fácil imaginar algún caso en el cual (hielo) expresa un pensamiento de agua; basta con que Sally infiera el pensamiento que expresaría profiriendo (hielo) exclusivamente de pensamientos de agua que guarda desde su estancia en la Tierra. Supongamos, por ejemplo, que Sally aprendió en la Tierra que el agua se hiela a cero grados. Un día, en la Tierra Gemela, recuerda algo que aprendió de pequeña, que expresaría profiriendo el enunciado (hielo). Ese pensamiento que tiene Sally en la Tierra Gemela es que el agua se hiela a cero grados. Está causalmente relacionado con un pensamiento original sobre agua y no hay presencia alguna del concepto BI-AGUA (ni de sus instancias). Por eso, sí es posible diseñar casos en los que Sally expresa un pensamiento de agua cuando profiere el enunciado (hielo) y, en esos casos, Sally recuerda semánticamente que el agua (no la bi-agua) se hiela a cero grados. Sally no pierde este conocimiento cuando transita a la Tierra Gemela.

Por contra, supongamos que Sally nunca aprendió en la Tierra que el agua se hiela a cero grados, y que nunca pensó un pensamiento tal. Supongamos que, una vez en la Tierra Gemela, un amigo suyo (quien carece del concepto AGUA, pues ha habitado toda su vida en la Tierra Gemela) le explica que la bi-agua se hiela a cero grados. Un día, en la Tierra Gemela, ante un estanque de bi-agua helada, Sally recuerda aquello que le enseñó su amigo: algo que expresaría profiriendo el enunciado (hielo). Hay al menos dos factores que pueden ser relevantes para determinar el contenido de este pensamiento: el recuerdo de Sally de que la bi-agua se hiela a cero grados y la presencia de un estanque de bi-agua helada. Por eso, creemos que en este caso el pensamiento que trae a la mente Sally, aquello que recuerda semánticamente, es que la bi-agua se hiela a cero grados, un pensamiento en parte constituido por el concepto BI-AGUA[175].

---

dudaríamos en adscribir un pensamiento de agua o de bi-agua, y casos en los que no sabríamos bien qué contestar. Esto es, no creemos que nuestras intuiciones al respecto sean confusas, tenemos la intuición de que la realidad es confusa, e intentamos que nuestra teoría capte esta idea.

[175] Gibbons (1996) propone un ejemplo de un sujeto que expresa pensamientos con conceptos distintos con el mismo enunciado, dependiendo de las prácticas lingüísticas de qué comunidad de hablantes esté siguiendo. Supongamos que John tiene dos primos que se llaman Vinnie (uno por parte de padre, el otro por parte de madre). Los dos primos se parecen mucho entre sí, y John erróneamente cree que tiene un

Por último, supongamos que Sally aprende en la Tierra que el agua se hiela a cero grados. Un día en la Tierra Gemela, Sally y sus amigos se encuentran con un estanque de bi-agua helada. Los amigos de Sally comienzan a discutir sobre si la bi-agua se hiela a cero o a cinco grados. En medio de la discusión, Sally recuerda aquello que aprendió en la Tierra, que el agua se hiela a cero grados, así como la vez en la que, en la Tierra, comprobó con una pequeña muestra de agua y una nevera que el agua se hiela a cero grados; así, profiere (hielo). Son varios aquí los factores relevantes para determinar el contenido del pensamiento de Sally (la presencia del estanque, la intención de Sally de participar en una práctica lingüística, sus recuerdos), y parece que tanto los conceptos AGUA y BI-AGUA como la bi-agua tienen una presencia importante en esos factores. Por eso, defendemos que no es de recibo interpretar el pensamiento de Sally ni como un pensamiento de agua ni como un pensamiento de bi-agua. La posición más plausible es, creemos, que el pensamiento de Sally está en parte constituido por un concepto amalgama.[176]

Pasemos ahora a considerar qué puede recordar y saber semánticamente Sally en t3, una vez ha descubierto que ha estado transitando de un escenario a otro, y que ha estado hablando y pensando sobre dos sustancias distintas entre las que no distinguía. ¿Puede Sally en t3 recordar semánticamente (y saber) que el agua se hiela a cero grados? No. Y, de nuevo, no es que Sally pierda este conocimiento porque ha perdido algún concepto, sino porque ha perdido justificación. A pesar de ello, a diferencia de lo que decíamos sobre la memoria episódica, Sally no ha perdido justificación mnemónica (recordemos que la memoria semántica no es un elemento en la justificación), sino de otro tipo.

único primo que se llama Vinnie. Un día, en una reunión de la familia de la madre de John, todos recuerdan cómo se comportó Vinnie (por parte de madre) en la boda de la tía Clara. Siguiendo la conversación de sus familiares, y recordando cómo bailaba Vinnie en la boda de la tía Clara, John profiere el enunciado 'Vinnie sabe bailar'. Otro día, en una reunión de la familia del padre de John, todos recuerdan cómo se comportó Vinnie (por parte de padre) en la boda del tío Tony. Siguienda la conversación de sus familiares, y recordando cómo bailaba Vinnie en la boda del tío Tony, John profiere el enunciado 'Vinnie sabe bailar'. Según Gibbons, estos dos enunciados expresan distintas creencias de John, una de ellas compuesta por el concepto VINNIE1 y la otra por el concepto VINNIE2. Es el hecho de que John está inmerso en prácticas lingüísticas distintas y el hecho de que se basa en memorias que refieren a "Vinnies" distintos lo que hace que esos dos pensamientos de John tengan contenidos distintos. Estamos completamente de acuerdo con Gibbons.

[176] Pongamos también un ejemplo de indeterminación. Hemos dicho que los casos de indeterminación parecen más comunes en los casos relacionados con los conceptos singulares. Supongamos que Peter cree, por un lado, que Pavarotti es italiano y, por el otro, que bi-Pavarotti tiene problemas de sobrepeso (es éste el ejemplo de transición lenta que estudiaremos en la tercera parte, ya lo describiremos entonces con más detalle). Debido a esas creencias que tiene, Peter adquiere una creencia que intentaría expresar profiriendo el enunciado 'Pavarotti es italiano y tiene problemas de sobrepeso'. Diríamos que en este caso no está claro cuál es el concepto que expresa Peter con Pavarotti, creemos que no está determinado cuál es el pensamiento que está pensando.

Supongamos que Sally aprende en la Tierra que el agua se hiela a cero grados. Cuando adopta esta creencia, además, tiene evidencia suficiente para que la creencia que ha adquirido constituya conocimiento. Pasan los años, Sally transita de una Tierra a la otra, y al final descubre cómo ha sido su pasado. Pero supongamos que Sally no adquiere evidencia nueva que pueda justificar su creencia de que el agua se hiela a cero grados (tampoco adquiere evidencia que pueda justificar su posible creencia de que la bi-agua se hiela a cero grados), y supongamos también que Sally no recuerda en qué contexto aprendió que el agua se hiela a cero grados. Como Sally no recuerda en qué contexto aprendió que el agua se hiela a cero grados, tampoco recuerda cuál era la evidencia que tenía a favor de su creencia original. Ahora, la memoria semántica de Sally funciona correctamente y, por eso, también en t3 le puede "traer a la mente" su creencia de que el agua se hiela a cero grados, en principio con su contenido y justificación originales. ¿Por qué no puede saber ahora Sally que el agua se hiela a cero grados? Defendemos que es la evidencia (no mnemónica) original que tenía Sally la que resulta minada ahora.

Sally tenía en t1 evidencia suficiente que justificaba su creencia de que el agua se hiela a cero grados (porque lo aprende en la escuela, o porque lo infiere de un experimento que realiza). Pero sucede que esa misma evidencia no podría justificar *ahora* la creencia de Sally de que el agua se hiela a cero grados. Cualquiera que fuera la evidencia que tuvo Sally para llegar a creer que el agua se hiela a cero grados, si ahora le presentáramos esa misma evidencia, no justificaría su creencia de que el agua se hiela a cero grados ya que, en cuanto Sally no sabe si esa evidencia es "evidencia de agua" o "evidencia de bi-agua", de esa evidencia Sally no puede inferir que el agua se hiela a cero grados (igualmente podría inferir que la bi-agua se hiela a cero grados porque, de acuerdo con todo lo que sabe Sally, podría ser "evidencia de bi-agua"). Para poder inferir que el agua se hiela a cero grados debería saber primero que es agua (y no bi-agua) eso que se ha helado cuando hemos puesto el termostato de la nevera a cero grados, o que su profesora está hablando sobre agua (y no bi-agua) cuando ha proferido el enunciado 'El agua se hiela a cero grado' Por eso, al descubrir que ha estado transitando de un escenario a otro, Sally ha adquirido información que influye en su situación epistémica, y pierde la justificación que tenía originariamente para creer que el agua se hiela a cero grados; su memoria semántica no puede preservar esa evidencia que tenía originariamente.

6. COMPARACIÓN CON OTRAS PROPUESTAS

En los últimos dos capítulos hemos esbozado una teoría sobre justificación mnemónica, y hemos presentado las predicciones a las que nos compromete. En el capítulo que sigue defenderemos que nuestra propuesta es preferible a aquéllas presentadas en el segundo capítulo. Como veremos, el modelo esbozado está comprometido a menor "perdida mnemónica" que la mayoría de los modelos basados en cohabitación (y, se supone, uno apuesta por la cohabitación conceptual porque quiere defender que no hay pérdida mnemónica), y, en comparación con el reemplazo conceptual, la cohabitación nos compromete a una caracterización más natural de la memoria.

Recordemos primero nuestras predicciones para las proferencias (a), (b) y (c), hechas por Sally en la Tierra Gemela y antes de descubrir que ha estado transitando:

(a) El agua se hiela a cero grados.
(b) En t1 pensé que el agua es a veces salada.
(c) En t1 el agua estaba fría.

Como hemos dicho, distintas proferencias de un enunciado que contenga el término 'agua' pueden expresar proposiciones distintas en boca de Sally. Dependiendo de qué factores tengan más presencia, las proferencias de (a), (b) y (c) pueden expresar tanto

proposiciones sobre agua como proposiciones sobre bi-agua. Ahora, si Sally se basa exclusivamente en su memoria y llega a un pensamiento que expresaría profiriendo alguno de los tres enunciados (a), (b) o (c), ese pensamiento estará en parte constituido por el concepto AGUA. Por eso, en ese caso Sally podrá recordar y saber que el agua se hiela a cero grados, o que en t1 pensó que el agua es a veces salada, o que en t1 el agua estaba fría.

6.1. COHABITACIÓN CONCEPTUAL Y PRESERVACIÓN DE CONTENIDO

Primero, no creemos que nuestra propuesta y la de Gibbons (1996) muestren grandes diferencias en cuanto a predicciones para (a), (b) y (c). Nuestro propósito era ofrecer un marco teórico según el cual no se seguiría que alguien perdería conocimiento mnemónico por el mero hecho de ser víctima de una transición lenta; siendo el lema de Gibbons que las transiciones lentas no cambian el contenido de los estados mentales guardados en la memoria, es fácil ver que llegaremos a las mismas conclusiones en cuanto a (a), (b) y (c). En los capítulos anteriores hemos mencionado la que creemos es nuestra mayor diferencia con Gibbons; éste se limita a mencionar que no hay cambio de contenido en los estados guardados en la memoria de Sally, pero no llega a proponer ninguna teoría de la justificación mnemónica. Nosotros hemos hecho explícito que creemos que la memoria episódica tiene un carácter *evidencial*, lo cual nos permitía defender que en t3 Sally no podía saber qué pensó en t1 sin estar comprometidos a defender que había perdido algún concepto. Nuestra propuesta sí se diferencia de la de Gibbons al menos en ese último punto. Además, a diferencia de Gibbons, nosotros defendemos que cuando Sally descubre que ha estado transitando de un escenario a otro no hay reemplazo conceptual.

Resumamos aquí brevemente las opiniones de Burge (1998) sobre estas cuestiones mnemónicas, que ya hemos presentado en el segundo capítulo de esta parte. Primero, Burge diferencia entre memoria preservativa y memoria sustancial. La memoria preservativa se encarga de "reactivar" las creencias que hemos adoptado anteriormente

(tiene la misma función de lo que nosotros llamamos 'memoria semántica'); la memoria sustancial, en cambio, nos proporciona recuerdos sobre episodios autobiográficos que hemos vivido (como la memoria episódica). Una de las diferencias más importantes entre memoria preservativa y memoria sustancial es que (y aunque suene a perogrullo) la memoria preservativa es preservativa, y la memoria sustancial no. Esto es así, se supone, porque la memoria preservativa se basa en cadenas causales (que van desde el pensamiento original hasta la recuperación mnemónica de ese mismo pensamiento) que posibilitan que se preserve el contenido del pensamiento en cuestión. Así, cuando alguien recuerda que la capital de Indonesia es Yakarta, la memoria preservativa preserva el contenido de la creencia original que recupera ahora y, por lo tanto, el sujeto que recuerda que la capital de Indonesia es Yakarta por ejemplo no necesita identificar de algún modo la creencia que recupera ahora. Por contra, cuando alguien recuerda sustancialmente que un día estuvo en Yakarta, sí ha de identificar aquello que recuerda y, por eso, en principio es posible que, transiciones lentas de por medio, confunda aquello que cree recordar con otras proposiciones.

Con esto en mente, mencionemos rápidamente cuáles serían las predicciones de Burge para las proferencias (a), (b) y (c) introducidas párrafos más arriba. Cuando Sally profiere (a) no está recordando un episodio autobiográfico; por lo tanto, es la memoria preservativa la encargada de que Sally pueda recordar aquello que expresa cuando profiere (a). Supongamos que Sally aprendió en la Tierra que el agua se hiela a cero grados y que, ahora, en t2, intenta recordar preservativamente aquel pensamiento original. Dado que hay una cadena causal entre el pensamiento original de que el agua se hiela a cero grados y el pensamiento que recuerda Sally en t2, parece que según Burge no hay mayores problemas para que recuerde que el agua (no la bi-agua) se hiela a cero grados. El pensamiento que expresa Sally con (a) está en parte constituido por el concepto AGUA y, dado que la memoria preservativa en la que se basa funciona correctamente, no hay más problemas para concluir que puede recordar y saber en t2 que el agua se hiela a cero grados. Además, Burge explícitamente dice que una proferencia del enunciado (a) podría expresar un pensamiento en parte constituido por el concepto BI-AGUA; basta con que no haya una cadena causal que una el pensamiento actual con un pensamiento original que esté en parte constituido por el concepto AGUA. Bien, por ahora nuestra propuesta no se diferencia demasiado de la de Burge.

Ya dijimos en el segundo capítulo que Burge defendía que algunas proferencias de (b) en boca de Sally en t2 podían expresar creencias de agua, y que, por eso, Sally podía recordar en t2 que en t1 pensó que el agua es a veces salada. Decíamos, Burge mantiene que Sally puede recordarlo así si su memoria preservativa y su memoria sustancial trabajan juntas. La memoria preservativa se encargaría sólo de guardar el contenido del pensamiento que tuvo Sally en t1, y la memoria sustancial de identificar el evento del pensamiento de Sally en su contexto. En cuanto a (c), Burge no dice nada acerca de casos de este tipo. Pero en principio parece que está comprometido a defender que las proferencias de (c) de Sally expresan creencias de bi-agua y que, por lo tanto, Sally no puede recordar en t2 que en t1 el agua estaba fría. Se trata de un ejemplo de memoria episódica, así que ha de ser la memoria sustancial de Burge la que se encarga de proporcionarle esos recuerdos a Sally, y ya hemos visto que según Burge este tipo de memoria no es preservativo. En t2 Sally cree recordar (erróneamente) que en t1 la bi-agua estaba fría.

Bien, pues sucede que tenemos serias dudas sobre las opiniones de Burge. Primero, como explicaremos en los párrafos siguientes, no vemos cómo puede defender, sobre la base del modelo que él ha esbozado, que en t2 Sally puede recordar que en t1 pensó que el agua es a veces salada; la falsa creencia de Burge de que sí puede hacerlo viene de que se basa, primero, en un uso ambiguo que hacen del término 'memoria semántica' algunos psicólogos como Tulving y, segundo, en textos demasiados antiguos como Tulving (1972). Como se verá, una vez dejamos de lado estos usos ambiguos, incluso sobre la base de textos más recientes de Tulving, no hay motivos para distinguir entre memorias que son preservativas y memorias que no lo son.

¿Cómo es eso de que la memoria preservativa y la sustancial trabajarían juntas en el caso de (b)? Como hemos visto, se supone que la memoria preservativa se encargaría de guardar el contenido del pensamiento de Sally en t1 y, así, ésta podría en t2 recordar que en t1 pensó que el agua es a veces salada (no que pensó que la bi-agua es a veces salada). Pero Burge no da ninguna explicación de por qué la memoria preservativa se dedicaría a guardar el contenido del pensamiento que tuvo Sally en t1 (y sólo de ese pensamiento); en principio parecería que la memoria preservativa es un mecanismo diseñado para "traer de nuevo a la mente" conocimiento proposicional (no autobiográfico) que una vez tuvo el sujeto, no se ve muy claramente por qué iba a tomar

parte en un ejemplo de memoria episódica. Burge (1998) sugiere (no dice explícitamente) que la distinción entre memoria preservativa y memoria sustancial (no-preservativa) se basa en la distinción hecha por algunos psicólogos entre memoria semántica y memoria episódica (Burge dice que "hay una analogía entre las dos distinciones"); nos remite a Tulving (1972) para la diferencia entre memoria semántica y memoria episódica. De hecho, si por "analogía" entendemos que la memoria preservativa *es* memoria semántica y que la memoria sustancial *es* memoria episódica, y si además aceptamos las ideas de Tulving (1972) sobre memoria semántica y episódica, parece más fácil comprender cómo se combinan la memoria preservativa y la sustancial en el caso de Sally y por qué la memoria preservativa se encarga de preservar el contenido del pensamiento que tuvo Sally en t1 (y sólo de eso).

Según Tulving (1972), mientras que la memoria episódica (encargada de guardar información acerca de nuestros episodios autobiográficos) guarda eventos perceptivos, la memoria semántica (encargada de guardar la información que tenemos acerca del mundo en general) es "el conocimiento organizado que una persona posee sobre palabras y otros símbolos verbales, su significado y referentes, sobre relaciones entre ellos, y sobre reglas, fórmulas, y algoritmos para la manipulación de estos símbolos, conceptos, y relaciones"[177]; además, la memoria episódica y la semántica pueden trabajar juntas. Siendo la memoria preservativa/semántica el mecanismo encargado de guardar nuestro conocimiento *semántico*, y siendo la memoria episódica la encargada de guardar información acerca de los eventos que hemos vivido, parece razonable pensar, como hace Burge, que en el caso de Sally, aunque la memoria sustancial/episódica se encarga de identificar como evento su pensamiento en t1, es la memoria preservativa/semántica la que guarda el contenido de ese pensamiento.

Pero esta lectura es errónea, y así lo sugieren textos más recientes de Tulving sobre memoria semántica y memoria episódica[178]. Primero, como ya hemos dicho, Tulving (1972) hace un uso ambiguo de los términos 'memoria semántica' y 'memoria episódica'; de hecho, Tulving (1990) distingue entre dos "sentidos" de 'memoria episódica'. Por un lado, 'memoria episódica' y 'memoria semántica' pueden referirse a

[177] ...organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of these symbols, concepts, and relations. (Tulving (1972), p. 386)
[178] Véanse, por ejemplo, Tulving (1990, 2001, 2002).

dos funciones diferentes que desempeña la memoria; por otro, a dos sistemas cognitivos distintos que posibilitan que la memoria cumpla esas dos funciones[179]. Así, habría que explicitar si la analogía de Burge es entre la memoria preservativa y la memoria semántica-(f) o entre la memoria preservativa y la memoria semántica-(c). Ninguna de las dos analogías funciona.

Si la analogía que defiende Burge es entre la memoria preservativa y la memoria semántica-(f), entonces no queda claro por qué motivo la memoria semántica-(f) debería de ser preservativa y la episódica-(f) no, no parece justificado asumir que una de esas funciones es preservativa y la otra no sin ningún motivo para ello. Además, si la analogía fuera ésta, no vemos cómo sería posible que la memoria preservativa y la sustancial "trabajaran juntas". La memoria semántica-(f) y la memoria episódica-(f) son dos funciones que ha de cumplir la memoria, realmente no se ve cómo podrían conjugarse. Por de pronto, no sabemos si el producto de esta conjunción mnemónica le aportará al sujeto conocimiento sobre sus propios episodios autobiográficos o conocimiento proposicional general.

Si la analogía que propone Burge es con los dos sistemas cognitivos que describe Tulving (1972) (lo cual parece más probable), creemos que sigue teniendo problemas. Tulving (2001) acentúa la necesidad de que estos sistemas cognitivos trabajen conjuntamente para cumplir las distintas funciones que pueda tener la memoria y subraya concretamente la importancia que tiene la memoria semántica-(c) en la memoria episódica-(f). A pesar de que es la memoria episódica-(c) la encargada de almacenar y extraer la información acerca de nuestros episodios autobiográficos, la memoria semántica-(c) se encarga de codificar esta información. Esto es, según textos más recientes de Tulving, la memoria semántica-(c), supuestamente la memoria que es "preservativa", no se encarga de codificar sólo los elementos "semánticos" (enunciado, términos, pensamientos, etc.) que aparecen en nuestros episodios autobiográficos y, además, no es su trabajo "guardar" esa información (la memoria episódica-(c) es la encargada tanto de almacenar como extraer esa información). No vemos en qué medida podría sugerir esto que la memoria preservativa es preservativa y la sustancial no.

[179] En los siguientes párrafos, en los que nos centraremos en Tulving y Burge, usaremos, para evitar ambigüedades, el término 'memoria episódica-(f)' para referirnos a la memoria episódica como función que ha de cumplir la memoria, y 'memoria episódica-(c)' para referirnos al proceso cognitivo que en parte posibilita que la memoria cumpla esa función.

Así, creemos que las ideas más recientes de Tulving sobre memoria semántica-(c/f) y memoria episódica-(c/f) casan mejor con nuestra propuesta que con la de Burge (1998). Primero, no vemos ningún motivo para defender que la memoria semántica-(c) es "preservativa" y la episódica-(c) no, o que la memoria semántica-(f) lo es pero la episódica-(f) no. De hecho, si la memoria episódica-(c) no fuera preservativa, dado que es la encargada de almacenar y sustraer la información en la base de la memoria episódica-(f), resultaría difícil para Burge el mantener que en t2 Sally sí puede saber qué pensó en t1; es la memoria episódica-(c) la encargada de almacenar la información autobiográfica cuando la transición de Sally acontece. Si tenemos que creer (como parece que lo cree Burge) que la memoria episódica-(c) no es preservativa, se sigue que toda la información autobiográfica de agua que tenía Sally mutará en información de bi-agua una vez la transición lenta acontece. Por eso, no creemos que sea necesario estipular en Filosofía de la Memoria una distinción entre mecanismos de la memoria basada en la "preservatividad" o no de éstas; la analogía entre memoria preservativa y memoria semántica-(c) y los textos más recientes de Tulving nos llevan a la conclusión de que toda la memoria semántica-(f) de Sally es preservativa, y toda la memoria episódica-(f) no preservativa, y realmente no vemos ningún motivo para querer defender ni la analogía ni su consecuencia.

No hay ningún motivo para distinguir entre memoria que es preservativa y memoria que no lo es. En principio, el defensor de la cohabitación de conceptos tiene la puerta abierta para asumir que toda memoria es preservativa. Por eso, creemos que nuestra propuesta es preferible a la de Burge (1998). Primero, a diferencia de Burge, nosotros sí podemos defender que Sally puede recordar y saber que en t1 el agua estaba fría. Además, creemos que nuestra propuesta casa mejor con las ideas más recientes de Tulving acerca de memoria semántica y memoria episódica.

6.2. REEMPLAZO CONCEPTUAL Y MEMORIA EPISÓDICA.

Las diferencias con aquellos que defienden que las transiciones son ejemplos de reemplazo conceptual son más notables. Según éstos, ningún estado mental de Sally posterior a la transición está en parte constituido por el concepto AGUA; de acuerdo con

Ludlow (1995b, 1996, 1999) estarían constituidos por el concepto BI-AGUA y, de acuerdo con Falvey (2003), por algún tipo de concepto amalgama. Veamos cuáles serían sus predicciones para (a), (b) y (c).

6.2.1. Predicciones

Según el modelo de Ludlow, con (a) Sally expresaría su creencia de que la bi-agua se hiela a cero grados, con (b) que en t1 pensó que la bi-agua es a veces salada, y con (c) que en t1 la bi-agua estaba fría; con (a) expresa una proposición que de hecho es verdadera (el agua y la bi-agua comparten sus propiedades *macro*: también la bi-agua se hiela a cero grados), con (b) y (c) proposiciones falsas. Según Falvey (2003), en cambio, con (a) Sally diría que el agua y la bi-agua se hielan a cero grados, con (b) que en t1 pensó que la dosaguas es salada (donde algo es dosaguas si es agua o bi-agua), y con (c) que algo que bien era agua o bien era bi-agua estaba fría en t1. Con (a) Sally expresa una proposición verdadera y con (b) una falsa, pero a diferencia de Ludlow, Falvey afirma que con (c) Sally dice algo que es verdadero.

Falvey y Ludlow están comprometidos a mantener que en t2 Sally no puede recordar y saber que en t1 pensó que el agua es a veces salada, todo teórico que abogue por el reemplazo conceptual ha de mantener, pues, que habrá algo que la víctima de la transición lenta no podrá recordar. Si hay reemplazo, Sally perderá todos los recuerdos de agua que tenía (aunque, en algunos casos, los reemplazará por recuerdos veraces de bi-agua o dosaguas).

En cuanto a la memoria semántica, en principio está claro que hay una parte de conocimiento que pierde Sally debido a la transición: Sally no puede recordar que el agua se hiela a cero grados (ya que ha perdido su antiguo concepto AGUA). Ahora, queda por averiguar si ha sustituido ese conocimiento que tenía por otro, queda por responder si la creencia de Sally de que la bi-agua[180] se hiela a cero grados constituye conocimiento o no. Según Ludlow, sí:

> Tales recuerdos, aunque posiblemente transitorios, no serían poco fiables como fuentes de conocimiento. Al contrario, no hay motivo alguno por el cual no puedan

[180] En los párrafos que siguen nos ceñiremos a Ludlow y sus ejemplos de BI-AGUA, pero todo lo que digamos concierne del mismo modo a Falvey y sus ejemplos de DOSAGUAS.

> ser completamente fiables en las condiciones contextuales en las que suceden. Por ejemplo, supongamos que en el momento t0 Sally aprende que el agua es húmeda. En el momento t1, antes de que cambie de entorno, puede recordar que el agua es húmeda. Después, en t2, debido a cambios no detectados en el entorno, puede tener el recuerdo de que la bi-agua es húmeda. ¿Es este segundo episodio de memoria menos fiable que la primera? Es difícil ver por qué.[181]

La creencia de Sally de que la bi-agua se hiela a cero grados es verdadera; es más, según Ludlow es una creencia fiable, no ve motivos para negar esto último. Uno podría defender que la creencia de Sally no es fiable del siguiente modo: aunque la creencia sea sobre bi-agua, se sustenta en aquello que aprendió sobre el agua; dado que el agua y la bi-agua son dos cosas distintas, podría ser que lo que es verdadero sobre el agua sea falso sobre la bi-agua—observar el comportamiento del agua para aprender sobre la bi-agua no puede ser un buen método. Pero Ludlow niega que este razonamiento sea convincente en los casos de transición lenta, ya que estos ejemplos implican la presencia de objetos o sustancias que han de compartir sus propiedades *macro*: "los casos de transición lenta son por definición aquellos casos en los cuales mi conocimiento personal de las condiciones de individuación del agua no es suficiente para distinguirlo de la bi-agua"[182]. Esto es, si Sally tiene la creencia verdadera de que el agua tiene la propiedad intrínseca P, entonces no se puede dar el caso de que, tras una transición lenta, llegue a creer falsamente que la bi-agua tiene la propiedad intrínseca P. Porque si esto fuera así, Sally conocería las condiciones de individuación del agua; la transición lenta no se podría dar completamente, Sally no perdería su concepto AGUA (y no adquiriría el concepto BI-AGUA) y seguiría creyendo que el agua tiene la propiedad P (no que la bi-agua tiene la propiedad P). Por eso, la memoria semántica de Sally después de la transición sí es fiable.

El modelo que hemos esbozado nosotros no tiene esas consecuencias de pérdida y reemplazo de conocimiento pero, ¿en qué medida supone eso una ventaja? Lo explicaremos en lo que queda de capítulo; intentaremos abogar por un modelo de cohabitación conceptual. Pero nos gustaría dejar claro desde el principio que no

[181] Such memories, although possibly transient, would not be unreliable as sources of knowledge. To the contrary, there is no reason at all why they cannot be completely reliable in those environmental conditions in which they occur. For example, suppose that at time t0 [Sally] come[s] to know that water is wet. At time t1, before [she] shift[s] environments, [she] may recall that water is wet. Later, at t2, due to undetected environmental changes, [she] may have the recollection that twater is wet. Is this second episode of memory less reliable than the first? It is difficult see why. (Ludlow (1996), p. 315)

[182] Slow switching cases are by definition those cases in which my personal knowledge of the individuating conditions of water is not sufficient to distinguish it from twater. (Ludlow (1996), p. 316)

creemos que haya ningún argumento que demuestre que uno de los dos modelos es claramente superior al otro; los dos modelos son coherentes. A pesar de ello creemos, primero, que en ausencia de argumentos que claramente favorezcan al reemplazo conceptual, el modelo basado en la cohabitación es preferible. Parece natural caracterizar a Sally en t2 como recordando que pensó que el agua es a veces salada o que el agua estaba fría; creemos que esta opción tiene cierta fuerza intuitiva. Pero, además, los dos modelos responden a diferentes opiniones acerca de cuál debe ser la función de la memoria, y creemos que algunas cuestiones al respecto mueven (aunque sólo sea ligeramente) la balanza a favor del modelo de cohabitación que hemos propuesto nosotros.

6.2.2. Reemplazo conceptual y memoria episódica.

Varios autores han criticado al modelo externista de la memoria que hace de la memoria una capacidad inútil. Al proporcionar creencias falsas, la memoria se vuelve una herramienta poco fiable en el modelo de Ludlow y, por lo tanto no puede aportar conocimiento—la memoria se vuelve fútil. Ya hemos visto cómo rechaza Ludlow estas críticas en cuanto se dirigen a la memoria semántica, centrémonos ahora en la memoria episódica. Entre otros, Ludlow (1996, 1999) y Bernecker (1998) responden que esas acusaciones presuponen un modelo de memoria que el externista no debería aceptar. Este modelo de la memoria presupuesto por el crítico de Ludlow asume que la función de la memoria es guardar nuestros pensamientos y creencias con su contenido original, cosa que Ludlow y Bernecker niegan. Y, a su modo, la memoria de Sally sí le proporciona información acerca de su pasado:

> Brevemente, para minar la interpretación externista de la memoria uno tiene que mostrar que hay una mayor ventaja epistemológica en una capacidad que podría guardar los contenidos de los episodios mentales iniciales en comparación con una capacidad que provee información a priori sobre episodios pasados, pero en relación con el entorno actual. Hoy por hoy no puedo imaginar cuál podría ser esa ventaja.[183]

> El problema con esta objeción a la noción de memoria de Ludlow es la asunción de que un estado mnemónico *necesariamente* contiene los contenidos y conceptos del

[183] In short, to undermine the externalist view of memory one has to show that there is a greater epistemological advantage to a faculty that could record the contents of initial mental episodes than there is to a faculty which provides a priori information about past episodes, but relative to current environmental conditions. I for one cannot imagine what the advantage could be. (Ludlow (1996), pp. 316-317)

> estado anterior relevante. (...) Lo que estoy sugiriendo entonces es que la función de la memoria, en vez de reactivar pensamientos guardados anteriormente, es proporcionar información sobre estados pasados en relación con las condiciones contextuales presentes. El transporte de contenidos y conceptos en el tiempo puede ser una condición suficiente para la memoria, pero no llega a ser una condición necesaria.[184]

Según Bernecker y Ludlow, pues, no hay por qué suponer que el quehacer de la memoria sea servir de almacén de nuestros pensamientos, parece plausible que el único objetivo de la memoria episódica sea proporcionarnos información acerca de episodios pasados. Y según ellos lo hace (aun y cuando hay reemplazo conceptual); lo hace "en relación con el entorno actual". Así, por ejemplo, el pseudo-recuerdo de Sally en t2 de que en t1 pensó que la bi-agua es a veces salada es un recuerdo *sobre* (no *de*) su vivencia en t1. La memoria le da a Sally cierta información sobre su vivencia en t1, la cual, aunque falsa, no tiene por qué no ser relevante o útil. En cuanto la memoria nos sigue dando información sobre nuestro pasado aun y cuando ha habido reemplazo conceptual, no se sigue que este modelo haga de la memoria un mecanismo fútil o inútil.

No creemos que esta respuesta sea aceptable. Primero, tal y como afirma Kraay (2002), no está claro qué es eso de que la pseudo-memoria de Sally en t2 sea *sobre* su pensamiento en t1:

> Están equivocados en cuanto a si "sobre" se ha de entender *intensionalmente* o *extensionalmente*, la "memoria" de t2 no puede ser *sobre* t1. Entendido intensionalmente, el pensamiento de bi-agua es sobre el concepto BI-AGUA, no sobre el concepto AGUA. Y entendido extensionalmente, el pensamiento de bi-agua es sobre el compuesto químico XYZ, el cual es claro, potable y cubre la mayor parte de la Tierra Gemela, no sobre el compuesto químico H_2O, el cual es claro, potable y cubre la mayor parte de la Tierra.[185]

[184] The problem with this objection to Ludlow's notion of memory is the assumption that a memory state *necessarily* contains the content and concepts of the relevant earlier state. (...) What I am suggesting then is that the job of memory, rather than to replay previously recorded contents, is to provide information about past states relative to the present environmental conditions. The transfer of contents and concepts across time might be a sufficient condition for memory but it falls short of being a necessary condition. (Bernecker (1998), p. 341)

[185] ...they are mistaken whether "about" is taken *intensionally* or *extensionally*, the t2 "memory" cannot be *about* [t1]. Taken intensionally, the twater-thought is about the concept *twater*, not the concept *water*. And taken extensionally, the twater-thought is about the chemical compound XYZ, which is clear, potable and covers most of the Twin Earth, and not H_2O, that is clear potable and covers most of the earth. (Kraay (2002), p. 301)

Parece que no hay nada en el contenido del recuerdo de Sally que haga referencia al pensamiento que tuvo en t1. Ludlow (1999) sí esboza lo que podría ser una explicación de cómo el recuerdo de Sally en t2 es un recuerdo de su pensamiento en t1. Según dice, el externista no tiene por qué admitir que las memorias han de ser individuadas sólo por sus contenidos externos, otros factores (como el rol funcional de ese estado mental, o su *contenido estrecho*) podrían ser relevantes a la hora de individuarlo. Siendo esto así, "uno puede todavía identificar los eventos mentales en t1 y t2 como del mismo tipo sobre la base de que tienen los mismos contenidos no-relacionales"[186]. Ya hemos aceptado antes que el rol funcional de un pensamiento puede ser relevante para su individuación, y que esta idea no está en contradicción con las tesis externistas. Ahora, lo que sí debe asumir el externista es que el contenido relacional de un estado mental es relevante para su individuación: si dos estados mentales tienen contenidos relacionales distintos, entonces esos dos estados mentales son dos estados distintos. El hecho de que el recuerdo de Sally en t2 tenga el mismo contenido estrecho (si es que hay algo así) que su estado mental en t1 no es suficiente para concluir que es un recuerdo sobre su pensamiento en t1; es más, el hecho de que ese recuerdo y el estado original de t1 tengan contenidos relacionales distintos es suficiente para que surjan dudas acerca de cómo es posible que el recuerdo sea *sobre* ese estado mental. Esta idea de Ludlow (1999) no es compatible con el externismo semántico.

Sea como fuere, creemos que Ludlow y Bernecker podrían responder a Kraay que el recuerdo de Sally es sobre su acto de pensar, sobre una vivencia, y que la relación causal entre la vivencia en cuestión y el recuerdo de Sally en t2 fija la referencia de este recuerdo. Así, el recuerdo de Sally en t2 diría falsamente sobre t1, o sobre el acto de pensar de Sally que aconteció en t1, que Sally pensó que la bi-agua es a veces salada—estarían comprometidos a negar que estas relaciones causales determinan en algún modo el contenido del recuerdo en t2, pero en principio esta posición parece coherente.

Pero aun y si aceptáramos esta explicación (que Ludlow y Bernecker no dan), es difícil ver cómo la pseudo-memoria de Sally en t2 le da *información* sobre t1. El recuerdo es a todas luces falso, no se entiende qué es eso de que la memoria le dé a Sally información en relación con su entorno actual. Por eso, puede que el reemplazo conceptual no haga

[186] ...one can still type-identify the t1 and t2 mental events by their having the same non-relational contents. (Ludlow (1999), p. 166)

de la memoria algo fútil, pero sí revienta su fiabilidad (por definición, una herramienta que sistemáticamente proporciona creencias falsas no es una herramienta fiable). Creemos que la memoria episódica sí supone un problema para el defensor del reemplazo conceptual. No sabemos si motivado por estas cuestiones, Ludlow (1999) prueba con otra alternativa:

> En general, si las creencias anteriores de Sally eran completamente fiables, también lo serán entonces sus creencias posteriores. Los casos en los que no se preserva la veracidad del recuerdo parecen suficientemente inofensivos. La concepción externista de la memoria que defiendo proporciona verdades que son importantes, y quizás proporciona algunas falsedades que no son importantes – en cualquier caso, no lo suficientemente importantes como para amenazar la supervivencia de Sally.[187]

Esto es, el objetivo de la memoria no es preservar contenidos, ni proveernos información acerca de nuestro pasado (en relación con nuestro presente), sino proveernos información que nos sea útil para desenvolvernos en nuestro entorno actual.

De nuevo, está claro que esta interpretación casa perfectamente con la memoria semántica, pero podría tener problemas con la memoria episódica. Parece que, por definición, ésta última nos proporciona información sobre nuestro pasado, no sobre nuestro presente. Pero, respondería Ludlow, aunque la memoria episódica nos proporcionara creencias falsas, dado que estas creencias emplean nuestro bagaje conceptual actual, nos son útiles para desenvolvernos en nuestro entorno actual (y se supone ahora que ésa es la función de la memoria). De esas creencias falsas el sujeto podría inferir creencias verdaderas sobre su entorno, creencias que, se supone, serían fiables aun y cuando se apoyaran en premisas falsas. La memoria es útil en el modelo de Ludlow (y no así, parece, en el modelo de cohabitación).

Primero, es falso que la memoria sea una herramienta "más útil" en el modelo externista de Ludlow que en un modelo basado en la cohabitación (como el propuesto por nosotros). Por un lado, porque no se sigue del modelo basado en cohabitación que alguien como Sally no tendría creencias de bi-agua, más adecuadas que las creencias de

[187] In general, if [Sally's] earlier beliefs were on the whole reliable, then [her] later beliefs will be as well. The failures of truth-preservation which we do encounter seem harmless enough. The externalist conception of memory I am advocating delivers truths that are important, and perhaps delivers some falsehoods that are not important – in any case, not important enough to undermine [Sally's] survival. (Ludlow (1999), p. 167)

agua (se supone) para desenvolverse en un entorno de bi-agua—el modelo de cohabitación es compatible con mantener que después de la transición Sally puede tener tanto creencias de agua como creencias de bi-agua (hemos visto que de acuerdo con nuestra propuesta Sally puede creer ambas cosas: que el agua se hiela a cero grados y que la bi-agua se hiela a cero grados). Por otro lado, incluso si negáramos que en un modelo de cohabitación Sally pudiera tener creencias de bi-agua que de algún modo se basaran en su memoria, no se seguiría que estas creencias no fueran útiles para ella en su entorno de bi-agua. Supongamos que Sally quiere preparar algunas bebidas, y que para ello pretende helar un poco de bi-agua que tiene. Recuerda que el agua se hiela a cero grados; basándose en esa creencia, y dado que confunde el agua con la bi-agua, pone el termostato de su nevera a -5º. Resulta difícil ver en qué medida un recuerdo de bi-agua sería más útil que este recuerdo de agua.[188]

No vemos cómo este modelo de memoria podría significar alguna ventaja para el reemplazo conceptual en comparación con la cohabitación. Por de pronto, el defensor de la cohabitación no tiene ningún problema para seguir con las caracterizaciones habituales de la memoria episódica y la memoria semántica. El modo más simple de caracterizar la memoria episódica es diciendo que es el complejo de mecanismos cognitivos que nos proporciona información (y conocimiento, se supone) sobre los diferentes eventos que hemos vivido, y tenemos serias dudas sobre si el teórico del reemplazo conceptual podría asumir esta definición.

Por eso, no creemos que el modelo de reemplazo muestre ninguna ventaja en comparación con el modelo de cohabitación. Además, a diferencia del teórico del reemplazo, el defensor de la cohabitación puede seguir con las definiciones habituales de memoria semántica y memoria episódica, y creemos que, aunque no un motivo aplastante a favor de la cohabitación, sí supone un punto a su favor.

6.2.3. Por qué no abogar por el reemplazo.

A lo largo de estas dos primeras partes del trabajo hemos mencionado, aquí y allá, algunas razones por las cuales la opción del reemplazo conceptual no nos parece del

[188] Alguien (Boghossian (1992a, 1994)) podría protestar que esta respuesta es problemática, porque se basa en una inferencia no válida que hace Sally. Dedicaremos toda la tercera parte a estas cuestiones.

todo convincente. En esta última sección de la segunda parte nos gustaría al menos recordar algunas de esas razones.

Primero, no parece que esté claro qué puede motivar que en una transición lenta haya reemplazo de conceptos, que la víctima de una transición pierda su concepto anterior por haber adquirido uno nuevo. Recordando a Gibbons (1996), "siendo fácil ver cómo un contacto causal con un nuevo tipo de sustancia puede proveerte con un nuevo concepto, no está del todo claro cómo te puede privar de uno".

Sally transita a la Tierra Gemela, y entra en contacto con una nueva sustancia, la bi-agua, y con una nueva comunidad lingüística, que usa el término 'agua' de un modo distinto—pero en ningún momento pierde "contacto cognitivo" con su entorno anterior, tiene memorias (o lo que sea) causadas por sus experiencias de agua. Siguiendo a Heal (1998), creemos que uno no puede simplemente obviar este contacto con un entorno anterior a la hora de individuar los conceptos que tiene Sally (y es exactamente esto lo que hace el defensor del reemplazo conceptual); Sally por ejemplo mencionaría instancias de agua como ejemplos de aquello que ella llama 'agua'.

Mencionamos en la primera parte que Heal (1998) parecía adherirse a un modelo que abogara por la adquisición de un concepto amalgama por parte de Sally: dado que ésta menciona tanto instancias de agua como instancias de bi-agua como instaurando el estándar del concepto que expresa mediante 'agua', parece que esto nos fuerza a adscribirle a Sally un concepto que abarque todas esas instancias.

Heal (1998) dice explícitamente que se limita al externismo de clases, y creemos que esta elección influye en su adopción de un modelo de reemplazo (en su vertiente de "concepto amalgama") en detrimento de un modelo de cohabitación. El nuevo entorno que habita Sally no difiere del anterior sólo en que contiene una clase natural que ésta carecía; también difiere en que la comunidad lingüística en la que toma parte Sally no es aquélla en la que participaba mientras estaba en la Tierra, y el externista social querrá decir que este cambio en el entorno de Sally es relevante a la hora de determinar qué conceptos tiene. La cuestión es que, si nos adherimos a las tesis típicas que defiende el externista social, querremos defender que Sally sí tiene el concepto BI-AGUA común en la nueva comunidad que habita (porque adquiere ese concepto *deferencialmente*, o

porque su intención de ceñirse al uso que hace de un término la comunidad lingüística de la que es parte nos lleva a adscribirle ese concepto). Además, como ya hemos dicho, aunque puede ser aceptable que alguien adquiera un concepto amalgama que refiere a una clase o a una propiedad, esta idea pierde plausibilidad para los casos en los que el concepto en cuestión es un concepto singular—no creemos que la idea de un concepto amalgama que refiera a Pavarotti y bi-Pavarotti sea una idea clara.

Porque Sally habita una nueva comunidad lingüística, queremos decir que tiene el concepto BI-AGUA. Porque Sally no ha perdido contacto con algunas instancias de agua, queremos decir que tiene algún concepto en parte determinado por aquellas instancias de agua. Opinamos que el mejor modo de incorporar estas dos ideas es apostando por un modelo de cohabitación—después de la transición Sally tiene los dos conceptos AGUA y BI-AGUA.

Por otro lado, el reemplazo conceptual proporciona resultados más bien extraños cuando se combina con otra tesis, la del predominio de las transiciones lentas (curiosamente, es el mismo Ludlow quien argumenta a favor de esta idea). Según Ludlow (1995a), el mundo actual contiene ejemplos de transición lenta, y estos casos no son tan extraños como en principio uno podría suponer:

> Cambiamos de una comunidad lingüística a otra frecuentemente y sin saberlo. Más concretamente, en cuanto viajamos entre diferentes círculos de conocidos, los contenidos de nuestras proferencias y nuestros pensamientos pueden cambiar del mismo modo.[189]
>
> Es común que nos movamos entre grupos e instituciones sociales, y en muchos casos no detectamos los cambios en el contenido de nuestros pensamientos.[190]

Hay varios términos que tienen dos usos distintos que no difieren demasiado entre sí[191]. En cuanto uno no sepa que hay dos usos distintos, no tenga conocimiento suficiente como para prevenir uno de los dos usos y use el término en cuestión deferencialmente, podrá ser víctima de una transición lenta. Ahora, si en el mundo actual hay individuos

[189] ...we frequently and unknowingly slide from one lanugage community to another. More to the point, as we travel between different circles of acquaintance, the contents of our utterances and thoughts may shift as well. (Ludlow (1995a), p. 227)
[190] We routinely move between social groups and institutions, and in many cases shifts in the content of our thoughts will not be detected by us. (Ludlow (1995a), p. 228)
[191] Ludlow pone ejemplos como 'chickory', 'football' o 'pragmatist'; en castellano podemos encontrar términos como 'carta', 'changurro', 'pelota', 'pragmático' o 'tenderete', con dos usos distintos relacionados, pero que no varían demasiado entre sí.

que sufren transiciones lentas y si éstas son ejemplos de reemplazo conceptual, se sigue que en el mundo actual hay individuos que no pueden recordar sus pensamientos pasados por haber sido víctimas de una de estas transiciones.

Pongamos un ejemplo. Marv no sabe mucho de filosofía y literatura, le interesan más el fútbol y los coches deportivos. Pero en su grupo de amigos todos son fanáticos de la lectura, y Marv muchas veces se siente algo desplazado, no entiende muchas de las discusiones en las que toman parte sus amigos. A pesar de ello, siempre intenta aprender algo de ellos, disimular su ignorancia. Una tarde en el bar donde se encuentran siempre, los amigos de Marv comienzan a discutir sobre las ventajas de una metafísica realista en comparación con una posición idealista; Marv (sinceramente) profiere el enunciado 'Hay motivos para intentar ser realista'. La tarde discurre entre cafés y tabaco, y los amigos de Marv cambian varias veces de conversación. Al llegar la noche, comienzan a discutir sobre sus inquietudes literarias, sobre realismo sucio y realismo mágico. De nuevo, intentando tomar parte en la conversación, Marv profiere (sinceramente) el enunciado 'Los realistas son gente interesante'.

Marv no sabe que las dos conversaciones tratan de gente distinta; los realistas sobre los cuales hablaban sus amigos al comienzo de la tarde no son los realistas de quien trataban al final de la noche. Y Marv ha adoptado *deferencialmente* los conceptos que forman las respectivas creencias—Marv tiene esas creencias, cree que hay que intentar ser realista en metafísica, y que los realistas (escritores) son gente interesante; ha proferido los enunciados sinceramente para expresar dos pensamientos que tenía en esos momentos. Sus dos usos del término 'realista' se basan en las convenciones que rigen los usos que hacen sus amigos (éstos son los "expertos" a los que *defiere* Marv). Por lo tanto, en principio podríamos decir que al comienzo de la tarde Marv cree que hay que intentar ser realista (metafísico), y al final de la noche que los realistas (escritores) son gente interesante; el término 'realista' expresa dos conceptos distintos en las dos proferencias de Marv, aunque éste no sea consciente de ello—Marv ha sido víctima de una transición lenta. Ahora, si éstas son ejemplos de reemplazo conceptual, se sigue que al final de la noche Marv carece del concepto REALISTA (METAFÍSICO), y que es incapaz de recordar que al comienzo de la tarde pensó que hay que intentar ser realista (en metafísica); cree falsamente que pensó que hay que intentar ser realista (en literatura).

Esto es extraño. Una vez aceptamos que hay cierto predominio de las transiciones lentas, abogar por el reemplazo conceptual resulta problemático. Se sigue que uno puede ser incapaz de recordar y repensar algo que pasó por su mente ese mismo día, o incluso durante la conversación que está siguiendo en ese momento. Esta consecuencia no es deseable.

Uno podría negar contra Ludlow (1995a) que haya predominio de las transiciones lentas; un modo evidente de hacer esto es aferrarse a la idea de que las transiciones lentas han de ser *lentas*, que uno necesita cierto tiempo para poder adquirir un concepto nuevo debido a un cambio inadvertido de entorno, y que por eso es imposible que a lo largo de una conversación (o una tarde, o un día, o una semana) uno pierda un concepto para adquirir otro.

Por supuesto uno puede adoptar esta posición; lo que sí parece claro es que en cuanto más importancia le demos a las intenciones de uno de ceñirse a los usos de su comunidad o a la deferencia a la hora de determinar qué conceptos tiene uno, más fuerza tendrá la idea de que hay predominio de las transiciones lentas—y menos la tesis del reemplazo conceptual.

Por último, a lo largo de este capitulo hemos defendido que la idea del reemplazo conceptual está reñida con cómo entendemos comúnmente qué es la memoria episódica. Es normal definir la memoria episódica como el conjunto de mecanismos cognitivos que nos proveen con información sobre los eventos que hemos vivido en primera persona. Ahora, si en las transiciones lentas hay reemplazo conceptual, entonces la información que recibirá la víctima de la transición sobre la base de su memoria episódica será en gran parte falsa. Parece que la memoria episódica de las víctimas de las transiciones lentas termina siendo poco fiable.

Resumiendo, no creemos que haya razones muy importantes que nos hagan elegir entre un modelo de cohabitación y otro de reemplazo conceptual—las dos posiciones son coherentes, y ninguna de ellas ha de enfrentarse con problemas demasiado graves. Pero sí creemos que hay algunos detalles que, aunque pequeños, favorecen la idea de que hay cohabitación y desde luego no vemos ninguna razón para abogar por el reemplazo conceptual. Aunque sólo sea por eso, nosotros apostamos, pues, por defender que

cuando uno sufre una transición lenta, termina con dos conceptos distintos entre los cuales no puede distinguir.

7. ÚLTIMOS COMENTARIOS Y CONCLUSIONES

En esta segunda parte hemos presentado un argumento incompatibilista, "el argumento de la memoria", que explota las supuestas consecuencias mnemónicas del externismo semántico, y nos hemos centrado en esas supuestas consecuencias. En el primer capítulo hemos presentado el argumento, en el segundo hemos defendido que no demuestra lo que pretende, ya que al menos una de sus premisas habrá de ser falsa.

Ahora, el argumento sí plantea cuestiones interesantes acerca de las consecuencias mnemónicas del externismo, y a estas cuestiones hemos dedicado toda la segunda parte.

Mencionemos las conclusiones más importantes a las que hemos llegado.

Primero, el argumento de la memoria no se sostiene, ya que se basa en un uso ambiguo de 'olvidar' (tal que si hiciera un uso no ambiguo del término, no podría ser que todas sus premisas fueran a la vez verdaderas). La primera premisa del argumento dice que si S sabe que *p* en t1, entonces, si S no olvida nada, S sabe que *p* en t2. La segunda premisa dice que Sally no olvida nada en el ejemplo que hemos presentado (algo que, se supone, podemos simplemente estipular).

La primera premisa propone una condición suficiente para que alguien olvide algo: que sepa que *p* en t1 pero que no en t2. El problema es que si aceptamos esta condición suficiente, entonces no podremos simplemente "estipular" en un ejemplo que alguien no olvida nada—si es consecuencia de las transiciones lentas que alguien pierde conocimiento, entonces uno no puede estipular al describir un ejemplo de este tipo que el sujeto "no olvida nada". Si el argumento hiciera un uso uniforme de 'olvidar', entonces alguna de las premisas sería falsa, éstas no pueden ser verdaderas al mismo tiempo. Por lo tanto no parece que el externismo semántico amenace nuestro auto-conocimiento autoritativo.

En el tercer capítulo hemos complicado un poco la historia de Sally, y hemos criticado las opiniones de Gibbons (1996) al respecto. Éste defendía que en el transcurso de t1 a t2 no había reemplazo conceptual, que Sally adquiría un concepto BI-AGUA pero mantenía su concepto AGUA. Pero sí había reemplazo según él en el transcurso de t2 a t3, cuando descubría cómo había sido su pasado, porque Sally perdía sus antiguos conceptos para adquirir ahora dos nuevos conceptos AGUA y BI-AGUA. Por eso, concluía Gibbons, Sally sí podía recordar en t2 qué pensó en t1, pero no así en t3.

Hemos respondido que estábamos de acuerdo con Gibbons en que en t3 Sally no podía recordar qué pensó en t1, pero que su argumento no era bueno. Primero, porque no era válido y, segundo, porque se basaba en una teoría errónea de individuación de conceptos. Así, hemos apostado por una teoría de conceptos "más externista" que afirme que Sally no adquiere ningún concepto nuevo al descubrir los detalles de su historia. Porque, si Gibbons estuviera en lo cierto, en t3 Sally no podría llegar a saber (ni siquiera empíricamente) qué pensó en t1—y esto es absurdo.

Hemos dicho que Sally no puede recordar en t3 qué pensó en t1, pero que esto no se debe a que haya perdido algún concepto que tenía en el primer momento. Así, nos hemos visto en la situación de explicar por qué afirmábamos que en t3 Sally no podía recordar, y hemos aducido que lo que pierde no eran conceptos, sino justificación epistémica. Para hacer explícitas cuáles eran estas relaciones justificatorias, hemos desarrollado lo que hemos venido a llamar 'una (proto)teoría de la memoria'. La teoría esbozada se basaba en la distinción entre memoria semántica y memoria episódica, aunque (a diferencia de Burge (1998)) asumía que las dos memorias preservaban el

contenido de aquello que se recordaba. Ahora, hemos defendido que cuando uno recuerda episódicamente que *p*, sí hay un elemento de naturaleza mnemónica en la justificación que tiene para creer que *p*, pero que no así cuando recuerda que *p* semánticamente—cuando uno recuerda episódicamente que *p*, lo hace sobre la base de una representación mnemónica que le sirve de evidencia; cuando recuerda semánticamente que *p*, la justificación de su creencia de que *p* no es más que la justificación que tuvo en el momento en el que adquirió la creencia originaria de que *p* que está en el origen causal de su recuerdo.

Después de esbozar esta teoría de la memoria hemos caracterizado las predicciones a las que está comprometida. Entre otras cosas, hemos dicho que Sally sí puede recordar en t2 qué pensó en t1, pero no así en t3 (porque adquiere evidencia que mina la justificación que tenía en t2).

En el último capítulo hemos comparado nuestra propuesta con aquellas que hemos presentado en el segundo capítulo. Entre otras cosas, hemos defendido que la propuesta de Burge (1998) no es plausible. Primero, nos parece confusa, no está claro cuál es la analogía entre memoria preservativa y memoria semántica que quiere hacer, y no se entiende bien qué quiere decir cuando afirma que hay casos en los que la memoria preservativa y la sustancial "trabajan juntas". Además, la analogía de Burge (sea cual sea ésta) no se sostiene; se sustenta en un uso ambiguo que hacen algunos psicólogos del término 'memoria semántica', y en textos demasiado antiguos. Si nos atenemos a algunos textos más recientes, vemos que no hay motivos para diferenciar entre memorias que son preservativas y memorias que no lo son.

Para terminar, hemos defendido que nuestra propuesta es preferible a aquellas que se basan en el reemplazo conceptual. Hemos dicho que las dos posiciones son en principio coherentes, que no hay argumentos de gran peso que favorezcan claramente una de ellas, pero también hemos expuesto tres motivos por los cuales sí pensamos que la opción de la cohabitación es, al menos, más atractiva que la del reemplazo. Primero, no vemos argumentos importantes para adoptar el modelo de reemplazo y, además, resulta difícil ver por qué el contacto con una nueva sustancia o una nueva comunidad lingüística tendría la consecuencia de que uno pierde algún concepto antiguo—cuando uno transita de un entorno a otro no pierde contacto cognitivo con el primer entorno, y

esto sugiere que ese primer entorno sigue siendo relevante pare determinar qué conceptos tiene aún después de la transición. Segundo, hemos defendido que el defensor del reemplazo tiene problemas para adoptar la caracterización tradicional de la memoria episódica. Normalmente definimos está como el conjunto de mecanismos cognitivos que nos permite recordar nuestras vivencias pasadas; ahora, si las transiciones lentas fueran ejemplos de reemplazo conceptual, la memoria episódica del sujeto le aportaría creencias que son de hecho falsas, y parece que esto reventaría la fiabilidad de la memoria. Por último, creemos que el reemplazo conceptual proporciona resultados más bien extraños cuando se combina con la idea (defendida por Ludlow) de que hay predominio de las transiciones lentas.

(3).....

EXTERNISMO, INFERENCIA Y RACIONALIDAD

0. INTRODUCCIÓN

Peter será el protagonista de nuestra última transición lenta. Supongamos que en t1 y antes de sufrir ninguna transición ni cambio, durante un viaje a Nueva Zelanda, Peter descubre al famoso tenor Luciano Pavarotti bañándose en el Lago Taupo. Gran amante de la ópera, esta visión le produce una gran emoción, es uno de los grandes momentos en la vida de Peter. Pasan los años, y varias veces recuerda que, una vez, se encontró con Pavarotti bañándose en el Lago Taupo. Al tiempo, sufre una transición lenta—un día se acuesta en la Tierra, pero (cosa que sucede inadvertidamente para él) amanece en la Tierra Gemela. Como ya hemos contado varias veces, la Tierra Gemela se parece mucho a la Tierra, aunque contiene bi-agua en vez de agua y, además, la Tierra Gemela tiene un curioso habitante, al cual llamaremos 'bi-Pavarotti'. bi-Pavarotti es muy parecido a Pavarotti: es italiano, canta ópera estupendamente, se llama 'Pavarotti' y tiene problemas de sobrepeso—la única diferencia es, podría decirse, que bi-Pavarotti no es Pavarotti. Pues bien, una vez ha pasado tiempo suficiente y Peter ha adquirido el concepto BI-PAVAROTTI (ha escuchado muchísimos de sus discos, y ha discutido sobre él muchas veces con sus amigos), lee en el periódico que bi-Pavarotti cantará en su ciudad. Emocionado, acude al recital. Al día siguiente, recuerda aquel viaje que realizó a Nueva Zelanda y al Lago Taupo, pensando un pensamiento que expresaría profiriendo el enunciado 'Una vez me encontré con Pavarotti bañándose en el Lago Taupo'; después, recuerda la noche anterior, pensando un pensamiento que expresaría

profiriendo el enunciado ‘Anoche escuché cantar a Pavarotti’. Finalmente, basándose en esas dos creencias, llega a la siguiente conclusión:

‘Un día me encontré en el Lago Taupo al tenor que escuché cantar anoche.’

Peter llega a esa conclusión mediante inferencia deductiva; es porque antes ha adoptado las otras dos creencias que llega a la conclusión que expresaría profiriendo el enunciado ‘Un día me encontré en el Lago Taupo al tenor que escuché cantar anoche’. Boghossian (1992a, 1994) se basa en este ejemplo para defender que el externismo semántico tiene consecuencias nefastas para nuestras adscripciones de racionalidad y actitudes proposicionales. Según él, si el externismo semántico fuera verdadero, la víctima de una transición lenta no podría evitar ser irracional, por mucho cuidado y esmero que pusiera en las inferencias y creencias que acepta—y, se supone, eso tiene consecuencias de lo más dañinas para nuestras adscripciones *de dicto*.

Como veremos, el argumento de Boghossian se basa en diversas premisas que defiende más o menos acertadamente. En esta tercera parte del trabajo presentaremos el argumento primero, haciendo explícita su estructura. Veremos luego que todos los pasos que da Boghossian son en principio discutibles y que, de hecho, todos ellos han sido atacados por un autor u otro. Algunos autores (Tye (1998), Schiffer (1992) o Burge (1998)) le critican que el externista no tiene por qué adscribir a Peter las creencias que dice él que está comprometido a adscribirle; otros simplemente aceptan que Peter sería irracional, pero entienden que esto no supone un gran problema para el externista. Nosotros defenderemos que la interpretación que hace Boghossian de la situación de Peter no es acertada, pero no porque Peter no tenga alguna de las creencias que le adscribe, sino porque Boghossian no tiene en cuenta en su explicación alguna creencia que sí tiene Peter. Por eso, defenderemos que el ejemplo no supone ningún tipo de problema para el externista: no hay puzzle.

1. TENORES E IRRACIONALIDADES

Boghossian (1992a, 1994) defiende que hay cierta tensión entre el externismo semántico y los motivos por los cuales adscribimos actitudes proposicionales *de dicto*. Es algo que diferencia las adscripciones *de re* de las adscripciones *de dicto* que cuando adscribimos un pensamiento o una creencia *de dicto*, entre otras cosas, lo hacemos para racionalizar la conducta y las creencias y juicios del sujeto. Según el argumento de Boghossian que presentaremos en este capítulo, el externismo semántico no es compatible con que nuestras adscripciones *de dicto* cumplan con esos cometidos. Y no lo es porque el externista está comprometido a negar que el contenido es transparente; uno de los objetivos de "Externalism and Inference" (1992) y de "The Transparency of Mental Content" (1994) es demostrar que "la tesis de la transparencia del contenido juega un papel importante en fijar qué es para nuestras actitudes proposicionales racionalizar nuestras conclusiones teóricas y prácticas"[192].

Boghossian (1992a, 1994) no da demasiados argumentos para justificar que el externista está comprometido a negar que el contenido es transparente; simplemente afirma que la víctima de una transición lenta estaría condenada a aceptar como verdaderos juicios de

[192] The thesis [of transparency of content] plays an important role in fixing what it is for our propositional attitudes to rationalize our practical and theoretical conclusions. (Boghossian (1992a), p. 26)

mismidad de contenido que de hecho son falsos[193]. Arguye, para apoyar esa tesis, que Peter por ejemplo tendría términos tipo de mentalés que, dependiendo del contexto en el que los ejemplificara, tendrían un contenido u otro.

Consideremos por ejemplo un pensamiento que Peter expresaría en la Tierra Gemela profiriendo el enunciado 'Este disco lo grabó Pavarotti' y un recuerdo de un episodio que vivió en la Tierra que expresaría profiriendo el enunciado 'De niño me pasaba el día escuchando discos de Pavarotti'. Boghossian asume que los dos pensamientos de Peter están en parte constituidos por dos ejemplares distintos del mismo término tipo de mentalés 'Pavarotti', que el primer pensamiento es claramente sobre bi-Pavarotti, estando por tanto en parte constituido por el concepto BI-PAVAROTTI y que, dado que son las vivencias de la Tierra las que fijan el contenido del recuerdo que tiene también en ese momento, el segundo pensamiento está constituido por el concepto PAVAROTTI. Pero Peter erróneamente pensaría que esos dos pensamientos son sobre lo mismo, que están constituidos por el mismo concepto. Por eso, la primera premisa del argumento de Boghossian sería:

(1) Si el externismo semántico es verdadero, entonces el contenido no es transparente.

Pero, al mismo tiempo, si la tesis de la transparencia es falsa, también se sigue que es falso que podamos conocer mediante pura reflexión *a priori* cuáles son las propiedades lógicas de nuestros pensamientos e inferencias:

> En sujetos como Peter, tanto la relación entre derivabilidad y validez como la transparencia de contenido de pensamientos caen, con el resultado de que inferencias que parecen válidas "desde dentro", no lo son. Y así, se muestra que la tesis de la *aprioricidad* de las capacidades lógicas resulta ser, por lo tanto, inconsistente con las asunciones externistas.[194]

[193] Esto en principio iría sólo en contra de la tesis de transparencia de diferencia de contenido. Boghossian asume que el externismo semántico también es incompatible con la tesis de la transparencia de mismidad de contenido, pero tenemos serias dudas de que cierto tipo de externismo neo-fregeano no sea compatible con esta tesis. Sea como sea, a Boghossian le basta con que el externismo esté reñido con la transparencia de diferencia de contenido para presentar su argumento. Por eso, también nosotros hablaremos de "transparencia del contenido" en general.

[194] In travellers like Peter, both the relationship between derivability and validity and the transparency of thought content break down, with the result that inferences that look to be "from the inside", valid, aren't. And the thesis of the *a priority* of logical abilities is shown, thereby, to be inconsistent, with externalist assumptions. (Boghossian (1992a), p. 22)

Esto es, si la reflexión *a priori* no nos garantiza la veracidad de nuestros juicios de mismidad y diferencia de contenido, entonces tampoco podremos saber *a priori* qué se sigue, y qué no, de alguno de nuestros pensamientos.

Peter llega a la conclusión de que una vez vio bañarse en el Lago Taupo al tenor que escuchó cantar anoche infiriéndolo de otras dos creencias que tiene; la primera de ellas es la que expresaría profiriendo el enunciado 'Una vez encontré a Pavarotti bañándose en el Lago Taupo', la segunda la que expresaría profiriendo 'Anoche escuché cantar a Pavarotti'. Si la primera premisa contiene el concepto PAVAROTTI y la segunda premisa contiene el concepto BI-PAVAROTTI (cosa que, según Boghossian, el externista está comprometido a mantener), se sigue que, aunque Peter cree que la inferencia es válida, de hecho no lo es (las premisas son de hecho verdaderas, y la conclusión de hecho falsa). Y la cuestión es que Peter erróneamente infiere la conclusión porque no puede ver que PAVAROTTI y BI-PAVAROTTI son dos conceptos distintos. Esto es, el externista está comprometido a asumir que alguien como Peter tendría creencias erróneas sobre qué se sigue y qué no de sus pensamientos y creencias:

(2) Si el contenido no es transparente, entonces uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias.

Queremos llamar la atención sobre un detalle aquí (ya que más adelante cuando presentemos la respuesta de Brown (2004) volveremos de nuevo a él). Lo que afirma Boghossian es que el externismo es incompatible con que, al menos para alguna inferencia concreta, podamos saber *a priori* si esa inferencia concreta es válida o no; no que el externismo es incompatible con que podamos saber *a priori* qué *formas* o *estructuras* argumentales son válidas y cuáles no:

> Evidentemente, no estoy diciendo que el externismo mina nuestra capacidad de decir *a priori* qué forma tendría que tener un argumento para ser lógicamente válido; estoy defendiendo que el externismo mina nuestra capacidad de decir *a priori* si una de nuestras inferencias en particular satisface alguna de esas formas.[195]

[195] Obviously, I am not saying that externalism undermines our ability to tell *a priori* what form an argument would have to have in order to be logically valid; I am arguing that externalism undermines our ability to tell *a priori* whether any particular inference of ours satisfies one of those forms. (Boghossian (1992a), p. 22, nota a pie de página 10)

La cuestión es que, para Boghossian, el que una teoría acerca de la naturaleza del contenido y nuestras adscripciones de actitudes proposicionales esté comprometida a negar que podemos conocer *a priori* las propiedades lógicas de nuestros pensamientos e inferencias tiene consecuencias nefastas para esa teoría. Se supone que cuando hacemos una adscripción *de dicto* intentamos racionalizar la conducta y las creencias del sujeto al que adscribimos la actitud:

> ¿Para qué necesitamos la tesis de la *aprioricidad* de la lógica? (...) una consideración intuitiva: la tesis juega un papel importante en fijar qué es para nuestras actitudes proposicionales racionalizar nuestras conclusiones prácticas y teóricas.[196]
>
> No adscribimos pensamientos a una persona sólo para decir algo descriptivamente verdadero acerca de ella. Usamos tales adscripciones para dos objetivos relacionados: por un lado, para poder aseverar su racionalidad y, por el otro, para explicar su conducta. (...) Manifestamos nuestra aceptación de este hecho al retirar los pensamientos *de re*—pensamientos que intuitivamente no son epistémicamente transparentes—de figurar en aseveraciones de racionalidad y explicaciones psicológicas.[197]

El problema es que, de acuerdo con Boghossian, negar que tenemos capacidades *a priori* como las descritas no es compatible con que nuestras adscripciones cumplan esos objetivos de racionalización.

Comencemos con las conclusiones teóricas—adscribimos pensamientos *de dicto* en parte para explicar la racionalidad del sujeto de la adscripción. Queremos decir que Peter es racional, hay un sentido en el cual lo es, querríamos que nuestras adscripciones *de dicto* explicaran esto. Y el problema es que, de acuerdo con Boghossian, las adscripciones a las que está comprometido el externista no pueden cumplir este cometido. Si la inferencia que piensa Peter tiene los contenidos que le adjudica el externista, entonces la inferencia no será válida, y Peter no podrá ver *a priori* que no lo es. Y, por lo que dice Boghossian, el poder identificar *a priori* los argumentos válidos es una condición necesaria para ser considerado racional:

---

[196] What do we need the thesis of the *a priority* of logic for? (...) an intuitive consideration: the thesis plays an important role in fixing what it is for our propositional attitudes to rationalize our practical and theoretical conclusions. (Boghossian (1992a), p. 26)

[197] We don't just ascribe thoughts to a person in order to say something descriptively true of him. We use such ascriptions for two related purposes: on the one hand, to enable assessments of his rationality and, on the other, to explain his behavior. (...) We manifest our recognition of this fact by barring *de re* thoughts—thoughts which intuitively lack epistemic transparency—from figuring in assessments of rationality and psychological explanation. (Boghossian (1994), p. 39)

> ¿Qué ha de hacer una persona para poder ser considerado un buen razonador? Claramente, en absoluto es una cuestión de conocer hechos empíricos, de tener montones de creencias verdaderas y justificadas sobre el mundo externo. Es más una cuestión de ser capaz de, y de estar dispuesto a, hacer que los pensamientos de uno se adecuen a los principios de la lógica sobre una base *a priori*. (...) Por lo tanto, la racionalidad es una función de las capacidades y disposiciones de una persona a adecuarse a las normas de la racionalidad sobre una base *a priori*; y las normas de la racionalidad son las normas de la lógica.[198]

La racionalidad tiene que ver con la capacidad para ajustar nuestras creencias a las normas que dicta la lógica *sobre la base de la mera reflexión a priori*. La víctima de una transición lenta como Peter no puede ajustar las creencias que le adscribe la teoría externista a las normas de la lógica sólo *a priori*; las adscripciones del externista no explican en qué sentido es Peter racional.

(C_1) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán explicar que es racional.

Esto es, la primera conclusión del argumento de Boghossian es que, si el externismo semántico es verdadero, entonces nuestras adscripciones *de dicto* no podrán racionalizar las creencias teóricas de los sujetos de las adscripciones, no pueden explicar que alguien como Peter es racional. Pasemos ahora a la explicación de la conducta.

Hemos citado a Boghossian afirmando que cuando adscribimos una creencia o un pensamiento no lo hacemos sólo para describir de cierta manera su estado mental, también lo hacemos, entre otras cosas, para explicar su conducta. Además, un sujeto será racional sólo si se comporta acorde a ciertas normas que relacionan conducta y creencias:

> Se espera de *cualquier* sujeto racional, no importa cuáles sean sus condiciones externas, que obedezca ciertas leyes: aquellas generalizaciones que reflejan las consecuencias lógicas introspectivamente obvias de las actitudes proposicionales de una persona. Así nuestra práctica psicológica ordinaria de explicar y predecir la conducta está construida sobre leyes como la siguiente:

[198] What does a person have to do in order to count a good reasoner? Clearly, it is not at all a question of knowing empirical facts, of having lots of justified true beliefs about the external world. Rather, it is a matter of being able, and of being disposed, to make one's thoughts conform to the principles of logic on an *a priori* basis. (...) So, rationality is a function of a person's ability and disposition to conform to the norms of rationality on an *a priori* basis; and the norms of rationality are the norms of logic. (Boghossian (1994), p. 42)

> Si **S** cree **p** [en t] y tiene [en t] la intención de **F** si **p**, y si S no tiene razones independientes para no acometer **F**, entonces S intentará acometer **F** o, al menos, tendrá la disposición de intentar acometer **F**.
> Si S tiene la intención de **F** syss **p**, pero no cree **p**, sino **q** en cambio, (donde **p** y **q** son proposiciones lógicamente independientes), entonces S no intentará acometer **F**.[199]

Cuando aún habita la Tierra y no ha transitado a la Tierra Gemela, Peter le manda por correo una bufanda de colores a Pavarotti. ¿Por qué actúa Peter de ese modo? Un modo de explicar su conducta es adscribiéndole el deseo de hacerle un bonito regalo a Pavarotti (F) y la creencia de que a Pavarotti le gustan las bufandas de colores (p). Pero ahora supongamos que Peter está en la Tierra Gemela, ya cuando ha adquirido el concepto BI-PAVAROTTI. Peter le manda por correo una bufanda de colores a bi-Pavarotti. ¿Por qué actúa así Peter? Parece que ya no tenemos a mano la misma explicación que teníamos antes. Tiene exactamente la misma creencia de que a Pavarotti le gustan las bufandas de colores, y el deseo de hacerle un bonito regalo a bi-Pavarotti, pero esto no explica su conducta. ¿Por qué, si lo que cree es que es Pavarotti a quien le gustan las bufandas de colores, le regala una a bi-Pavarotti? Los principios mencionados arriba no son buenos instrumentos para explicar la conducta cuando se combinan con una semántica externista.

(C_2) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán racionalizar su conducta.

Resumiendo, he aquí el argumento que presenta Boghossian (1992a, 1994):

(1) Si el externismo semántico es verdadero, entonces el contenido no es transparente.
(2) Si el contenido no es transparente, entonces uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias.

[199] *Any* rational subject, regardless of his external conditions, may be expected to obey certain laws: namely, those generalizations that mirror the introspectively obvious logical consequences of a person's propositional attitudes. Thus, our ordinary psychological practice of explaining and predicting behavior is built upon appeal to such laws as this:

> If S occurrently believes **p** and occurrently intends to **F** if **p**, and if S has no independent reason for not **F**'ing, then S will intend to **F** or, at the very least, will be disposed to intend to **F**.
> If S intends to **F** iff **p**, but does not believe **p**, but merely **q** instead, (where **p** and **q** are logically independent propositions), then S will not intend to **F**. (Boghossian (1994), pp. 42-43)

(C_1) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán explicar que es racional.

(C_2) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán racionalizar su conducta.

Mencionemos, aunque sea de paso, que Boghossian cree que las conclusiones del argumento en cierto modo minan nuestro auto-conocimiento autoritativo. Parece que para él el auto-conocimiento no se limita a saber qué estamos pensando, qué pensamos anteriormente, o qué creemos o deseamos, también incluye saber las relaciones de mismidad y diferencia de los contenidos de nuestros pensamientos:

> En un sentido importante, el externismo es inconsistente con la tesis de que tenemos conocimiento autoritativo en primera persona de los contenidos de nuestros pensamientos: en particular, defenderé, es inconsistente con la tesis de que los contenidos de nuestros pensamientos nos son *epistémicamente transparentes*.[200]

> El blindaje que nos proporciona este tipo de propuesta [Heil (1988), Davidson (1987), Burge (1988)] sobre la compatibilidad del externismo con el auto-conocimiento autoritativo es, en cierto sentido, *vacuo*: no conlleva las *consecuencias* típicas de la autoridad en primera persona sobre contenido de pensamientos.[201]

Bueno, ya hemos dicho suficiente sobre la supuesta tensión entre el externismo semántico y el auto-conocimiento autoritativo en la primera parte de este trabajo. Como dijimos, creemos que el externista debería rechazar la tesis de la transparencia del contenido, pero no vemos por qué esto debería amenazar el auto-conocimiento autoritativo de nadie.

Concluyamos este capítulo. ¿Qué alternativa le queda al externista según Boghossian? Parece que, o bien abandona la idea de que somos racionales, o bien ofrece una nueva noción de racionalidad:

---

[200] ...in an important sense, externalism is inconsistent with the thesis that we have authoritative first-person knowledge of thought content: in particular, I shall argue, it is inconsistent with the thesis that our thought contents are *epistemically transparent* to us. (Boghossian (1992a) p. 13)

[201] ...the assurance that this sort of proposal [Heil (1987), Davidson (1987), Burge (1988)] provides, about the compatibility of externalism with authoritative self-knowledge, is, in a sense to be explained, *hollow*: it carries with it none of the usual *consequences* of first-person authority about thought content. (Boghossian (1992a), p. 15)

> Si abandonamos la transparencia incluso para los pensamientos *de dicto*, (...) entonces debemos o bien abandonar la noción de racionalidad y con ella la práctica de explicación psicológica que sustenta, o bien debemos demostrar que estas nociones pueden ser remodeladas de modo que no lleven a resultados absurdos. El problema es que la primera sugerencia es descabellada y no parece que haya un modo obviamente satisfactorio de implementar la segunda.[202]

Bueno, de hecho hay quien defiende que necesitamos de evidencia empírica *a posteriori* para ser racionales[203], y otros recomiendan que "remodelemos" nuestras nociones tradicionales de racionalidad[204]. Además, uno también podría intentar rechazar alguna de las premisas del argumento, o la interpretación que hace Boghossian de la situación en la que se encuentra Peter.

De todas estas cuestiones trataremos en esta última parte del trabajo.

[202] ...if we abandon transparency even for *de dicto* thoughts, (...) then we must either jettison the notion of rationality and with it the practice of psychological explanation that it underwrites, or we must show these notions can be refashioned so as not to yield absurd results. The problem is that the first suggestion is wild and there appears to be no obviously satisfactory way of implementing the second (Boghossian (1994), pp. 39-40)
[203] Sorensen (1998), Williamson (2000), Faria (2009).
[204] Salmon (1986, 1989), Brown.

2. REEMPLAZO, TRANSPARENCIA, ANÁFORA

Una posibilidad de resistencia al argumento es rechazar alguna de sus premisas. En este capítulo presentaremos dos estrategias que optan por esta vía y que tienen en común que niegan que en el argumento de Peter tomen parte los dos conceptos PAVAROTTI y BI-PAVAROTTI. La primera estrategia que describiremos se basa en el reemplazo conceptual; negando que el ejemplo de Peter es un caso de cohabitación de conceptos uno puede aferrarse a la idea de que el contenido es transparente y, así, rechazar la primera premisa del argumento de Boghossian. Más tarde nos centraremos en la que seguramente es la respuesta más conocida y comentada, la defendida por Schiffer (1992) y Burge (1998). Ninguna de estas dos estrategias nos convence.

2.1. REEMPLAZO CONCEPTUAL Y TRANSPARENCIA DEL CONTENIDO

Una de las opciones que tiene a mano el externista es rechazar la premisa (1) del argumento de Boghossian, a saber, que el externismo es incompatible con la transparencia del contenido. Si Peter no confunde su concepto PAVAROTTI con su concepto BI-PAVAROTTI, entonces parece que podrá saber mediante la mera reflexión *a*

priori qué se sigue (y qué no) de sus pensamientos de 'Pavarotti' y que, por lo tanto, podrá conocer *a priori* las propiedades lógicas de sus pensamientos. La vía más fácil (si no la única) que tiene el externista para defender que el externismo y la transparencia son compatibles es sumarse a la idea de que en las transiciones lentas hay reemplazo conceptual. De hecho, Tye (1998) se basa en un escenario parecido al de Peter para argumentar en favor del reemplazo y afirmar que el externismo y la transparencia son compatibles[205]:

> 'Ahora sólo bebo agua antes de las cinco de la tarde. Años atrás bebía agua mezclada con ginebra en la sobremesa. Disfrutaba esas sobremesas—el agua mejora si se mezcla con ginebra.' Supongamos que, aunque yo no lo sepa, estoy ahora en la Tierra Gemela, y que llevo allí ya el tiempo suficiente, tal que he establecido en mi entorno las conexiones necesarias para los nuevos pensamientos y creencias. En estas circunstancias, mi término 'agua', usado como en el primer enunciado de mi alocución, significa bi-agua. (...) En el segundo enunciado, de nuevo lo más plausible es pensar que 'agua' significa bi-agua aún y cuando lo manifestado es sobre el pasado. Porque si 'agua' significa aquí agua, entonces yo no estoy *comparando* el que yo actualmente beba bi-agua con que *lo* bebiera en el pasado. Y eso parece erróneo. Intuitivamente, el concepto que empleo cuando uso 'agua' en el segundo enunciado refiere a la bi-agua, la misma sustancia sobre la cual dirigí mis comentarios del comienzo. (...) En el tercer enunciado, explico por qué disfrutaba esas sobremesas lejanas al añadir que la bi-agua mejora cuando se mezcla con ginebra. Esto no tendría sentido si lo que realmente creo es que bebí *agua* con ginebra durante esas sobremesas.[206]

Heal (1998) ofrece una esquematización muy clara del argumento de Tye. Según ella, el argumento se reduce a estas dos premisas:

(a) Tye expresa el mismo concepto cuando profiere el término 'agua' en las cuatro premisas de su alocución, también en sus recuerdos de la Tierra.

---

[205] Tye (1998) no trata el argumento que estamos presentando en esta parte, sino cuestiones relacionadas con el externismo y la memoria. Por eso, la posición que describiremos más adelante no es tanto la posición adoptada por Tye, sino una hipotética respuesta basada en sus opiniones (aunque no creemos que las ideas de Tye sobre el argumento de esta parte difieran mucho de las que caracterizamos).

[206] 'Water is the only thing I now drink before 5pm. Many years ago, however, I drank water fortified by gin in the afternoons. I enjoyed those afternoons—water is improved by mixing with gin.' Suppose that, unknown to me, I am now on Twin-Earth, and that I have been for a long time so that I have established the environmental connections needed for new thoughts and beliefs. In these circumstances, my word 'water', as it is used in the first sentence of my report, means twater. (...) In the second sentence, 'water' again is most plausibly taken to mean twater even though the report concerns the past. For if 'water' here means water, then I am not *comparing* my present drinking of twater with my past drinking of *it*. And that seems plainly wrong. Intuitively, the concept I exercise when I use 'water' in the second sentence refers to twater, the very stuff upon which I directed my opening remark. Intuitively, the concept I exercise when I use 'water' in the second sentence refers to twater, the very stuff upon which I directed my opening remark. (...) In the third sentence, I explain why I enjoyed those distant afternoons by adding that twater is improved by mixing it with gin. This would make no sense if what I really believed is that I drank *water* with gin during those afternoons. (Tye (1998), p. 81)

(b) Tye expresa el concepto BI-AGUA al menos en alguna de esas proferencias.

Se sigue por supuesto que todas las proferencias de 'agua' en la alocución expresan el concepto BI-AGUA. La explicación más simple es, según Tye, que después de la transición ha perdido el concepto AGUA que tenía, que éste ha sido reemplazado por el concepto BI-AGUA.

Heal (1998) acepta (a) y rechaza (b); nosotros sí creemos que hay proferencias de 'agua' en la alocución de Tye que claramente expresan el concepto BI-AGUA, y tenemos más dudas sobre (a)—pero dejaremos estas cuestiones de lado por ahora. Queremos mencionar qué razones da Tye en favor de (a), la tesis de que todas las proferencias de 'agua' en su alocución expresan el mismo concepto. Según él, si esto no fuera así, no podría *comparar* sus sobremesas actuales con sus sobremesas pasadas, algo que parece que es evidente que está haciendo, ni tampoco podría su creencia de que la bi-agua mejora cuando se mezcla con ginebra *explicar* que disfrutara de sus sobremesas pasadas.

Evidentemente, el que un pensamiento *explique* otro o que alguien *compare* entre sí dos pensamientos o hechos está estrechamente relacionado con que uno conozca las propiedades lógicas de sus pensamientos. Al fin y al cabo, su creencia de que la bi-agua mejora con la ginebra *explica* que disfrutara de sus sobremesas pasadas porque de su creencia de que disfrutaba esas sobremesas *se sigue* su creencia de que la bi-agua mejora con la ginebra. Porque de hecho Tye está explicando algunas de sus creencias mediante otras, se sigue según Tye (1998) que comparten conceptos (concretamente, comparten el concepto BI-AGUA). Porque Tye carece del concepto AGUA y lo ha reemplazado por el concepto BI-AGUA, se sigue que puede conocer *a priori* las propiedades lógicas de sus pensamientos constituidos por el concepto BI-AGUA.

Vayamos un poquito más despacio y esbocemos brevemente lo que sería la estrategia basada en el reemplazo conceptual, inspirada en Tye (1998). Si los ejemplos de transición lenta son casos de reemplazo conceptual, entonces parece que no hay motivos para pensar que el externismo semántico y la transparencia del contenido son incompatibles. Una vez ha adquirido el concepto BI-PAVAROTTI y ha perdido el concepto PAVAROTTI, no hay dos pensamientos que pueda tener Peter en ese momento

sobre los cuales cree erróneamente que tienen el mismo contenido. ¿Cuáles podrían ser los dos conceptos que confunde Peter? No pueden ser los conceptos PAVAROTTI y BI-PAVAROTTI, ya que Peter carece del primero. Volviendo de nuevo a Tye (1998), afirma que cree que el contenido es transparente, y que sus afirmaciones son compatibles con esa tesis:

> Es una tesis prima facie muy plausible, que una persona puede siempre saber autoritativa y directamente con respecto a cualesquiera de dos de sus pensamientos o creencias *presentes* si tienen los mismos contenidos o no. Opino que esta tesis comparativa, la cual es obviamente un pariente muy cercano de la tesis del acceso privilegiado para contenidos de pensamientos, no es tal que el externista tenga necesidad alguna de negar.[207]

Del mismo modo, ya que el externista puede mantener que el contenido es transparente, también puede afirmar que alguien en la situación de Peter podría conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias. ¿Cuál sería el argumento no válido que aceptara como válido? Si en el ejemplo que hemos expuesto al comienzo del capítulo Peter ha perdido el concepto PAVAROTTI, entonces de la premisa de que una vez se encontró con bi-Pavarotti en el lago Taupo y la premisa de que anoche escuchó cantar a bi-Pavarotti se sigue la conclusión de que una vez se encontró en el Lago Taupo con el tenor que escuchó cantar anoche. La conclusión es falsa, pero porque así lo es también una de las premisas, que Peter se encontró con bi-Pavarotti en el Lago Taupo; ahora, el argumento sí es válido. En principio no parece que Peter esté en peligro de aceptar como válido ningún argumento que de hecho no lo es.

Y, por eso, porque no hay ningún argumento que Peter acepta erróneamente como válido, no se sigue que las adscripciones del externista no puedan racionalizar las creencias y la conducta de Peter. Primero, Peter puede saber *a priori* qué se sigue y qué no de cada uno de sus pensamientos, puede adecuar *a priori* sus pensamientos a las normas de la lógica. Segundo, esas adscripciones explican perfectamente su conducta: manda una bufanda a bi-Pavarotti porque ahora cree que a bi-Pavarotti (no a Pavarotti) le gustan las bufandas de colores.

---

[207] [It is] prima facie a very plausible thesis, namely that a person can always know authoritatively and directly with respect to any two of his *present* thoughts or beliefs whether or not they have the same contents. In my view, this comparative thesis, which is obviously a close cousin of the privileged access thesis for individual thought contents, is not one that the externalist has any need to deny. (Tye (1998), p. 84)

2.1.1. Problemas con el reemplazo conceptual.

Esta estrategia, y sin considerar otros razones independientes que nos hagan dudar sobre el reemplazo conceptual, sí supone una buena respuesta al argumento de Boghossian. Alguien que apueste por el reemplazo puede aferrarse a la transparencia del contenido, y así rechazar la primera premisa del argumento. Pero la estrategia no nos convence. Primero, queremos mencionar que los argumentos de Tye a favor del reemplazo no nos parecen del todo convincentes y, además, ya hemos expuesto más de una vez nuestras dudas acerca de la idea de que alguien pierda el concepto AGUA cuando adquiere el concepto BI-AGUA. En esta sección nos limitaremos a mencionar esos motivos que ya hemos explicado antes.

El argumento de Tye se basa en las premisas (a) y (b). Primero, alguien podría rechazar (b), como lo hace Heal (1998). Según ésta, la víctima del reemplazo seguramente mencionaría ejemplares de agua como formando el *estándar* del concepto que expresa profiriendo el término 'agua', de lo cual concluye que la víctima no adquiriría el concepto BI-AGUA (como asume Tye con (b)), sino que reemplazaría su concepto AGUA por algún tipo de concepto amalgama. Nosotros en cambio tenemos dudas sobre la premisa (a).

El único argumento que da Tye a favor de (a) es que, en su alocución, está explicando uno de sus pensamientos mediante otro, y esto no podría ser así si no compartieran algún concepto. Esto no es convincente:

> Podemos explicar cómo es que el sujeto juzga que está comparando sus relaciones pasadas con una sustancia particular con sus relaciones presentes con ella, aún y cuando no lo está haciendo así. Al fin y al cabo, sistemáticamente confunde el agua con la bi-agua. Si uno reemplazara la última noche la impresora de mi ordenador por otra similar, hoy podría juzgar que estoy comparando el zumbido que hace al imprimir con el sonido que hacía ayer. De hecho, no podría estar haciendo esta comparación, pero es comprensible que juzgue que lo estoy haciendo.[208]

[208] We can make perfect sense of the subject's taking himself to be comparing his past relations with a particular substance to his present relations to it, even though he is not in fact doing that. He systematically mistakes water for twater, after all. If someone replaced my computer printer last night with a look-alike, today I might take myself to be comparing the whirring noise it makes while printing with the noise it made yesterday. I could not actually be doing so, but my taking myself to be doing this is quite comprehensible. (Collins (2008), p. 558)

Esto es, si uno cree que *a* es *b*, entonces si cree que *Fa* explica *Ga*, también creerá que *Fb* explica *Ga*. Idoia cree que Rosa es Maider, y nos pregunta por qué no está aquí Maider. 'Rosa está en clase de trompeta', le respondemos. ¿Explica el que Rosa está en clase de trompeta que Maider no está aquí? Depende, si Rosa es Maider, entonces sí. Independientemente de que sea una buena explicación, ¿cree Idoia que el que Rosa esté en clase de trompeta explica por qué Maider no está aquí? Sí.

Por eso el argumento de Tye no es bueno, del hecho de que crea que la creencia que expresa al proferir 'El agua mejora cuando se mezcla con ginebra' explica la creencia que expresa al proferir 'Disfrutaba las sobremesas en las que bebía agua con ginebra' no se sigue que esas dos creencias comparten conceptos; lo único que sigue es que él cree que hay una relación explicatoria (y que cree que comparten conceptos).

Y, como hemos dicho, tenemos serias dudas sobre la plausibilidad del reemplazo conceptual. Ya las hemos explicado en capítulos anteriores, no las desarrollaremos de nuevo aquí—pero sí nos gustaría enumerarlas de nuevo:

- Creemos que combinarla con la idea del predominio de las transiciones lentas (la cual juzgamos verdadera) da resultados poco deseables.
- No vemos por qué el encuentro con una nueva clase natural (un nuevo objeto o una nueva comunidad de hablantes) debería resultar en la pérdida de un concepto anterior.
- Tenemos dudas de que el defensor del reemplazo pueda ofrecer un modelo plausible de memoria episódica.
- Siguiendo a Heal (1998), el que Tye mencionara instancias de agua como ejemplificaciones de aquello que llama 'agua' es indicio de que esas instancias en parte fijan el estándar del concepto que expresa al proferir 'agua'. Esto dificulta asumir que Tye ha perdido su concepto AGUA.

2.2. REFERENCIA ANAFÓRICA Y HERENCIA CONCEPTUAL

Schiffer (1992) y Burge (1998) proponen una respuesta diferente al argumento, compatible con la cohabitación de conceptos. Brevemente, defienden que las

subsiguientes premisas y la conclusión del argumento heredan los conceptos de la primera premisa:

> ¿Cómo deberíamos entender la aparición del término en la segunda premisa? Sin lugar a dudas, *del mismo modo en el que se entiende en la primera premisa, cualquiera que sea ese modo*. Sea cual sea el modo de presentación operativo para Peter en la primera premisa, también será operativo para él en la segunda: y es que su intención es hablar acerca de la misma persona.[209]

> Peter dirá lo mismo mediante las dos proferencias de 'Pavarotti'. No veo otro modo de explicar la disposición de Peter –la cual sin lugar a dudas está implícita en cómo entendemos intuitivamente el ejemplo—de reformular la segunda premisa del siguiente modo: Pavarotti, quien una vez nadó en el lago Taupo, es el tenor que escuché ayer.[210]

Por lo tanto, niegan que el externista está comprometido a mantener que la víctima de una transición lenta no podría conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, que estaría en peligro de aceptar como válidos argumentos que en realidad no lo son (sin ser capaz de corregir su error mediante la sola reflexión *a priori*) [211]. Volvamos al ejemplo de Peter. La primera premisa de su argumento, que

---

[209] How should we understand [the term's] occurrence in the second premise? Surely, *in the same way it's understood in the first premise, whatever that way is*. Whatever mode of presentation is operative for [Peter] in the first utterance will also be operative with him in the second: for his intention is to be speaking about the same person. (Schiffer (1992), p. 33)

[210] Peter will mean the same by both tokens of 'Pavarotti'. I see no other way to explain Peter's willingness –which is surely implicit in our intuitive understanding of the example—to restate the second premise thus: Pavarotti, who once swam in Lake Taupo, is the singer I heard yesterday. (Schiffer (1992), p. 34)

[211] Tenemos dudas sobre si las propuestas de Schiffer (1992) y Burge (1998) van en contra de la premisa (1) o la premisa (2) del argumento de Boghossian; tenemos dudas sobre si creen que el externismo es compatible con la transparencia del contenido o no. Tendemos a pensar que aceptarían (1) y rechazarían (2), que negarían que el contenido es transparente; así nos hace pensar el hecho de que crean que Peter puede tener tanto pensamientos constituidos por el concepto BI-PAVAROTTI como recuerdos constituidos por el concepto PAVAROTTI aún y cuando no sabe que esos conceptos refieren a dos individuos distintos. Además, a pesar de que en ninguno de sus textos emplea el término 'transparencia' al referirse a estas cuestiones, Burge (1988) por ejemplo si que habla de la incapacidad de la víctima de una transición lenta de *discriminar* entre el pensamiento concreto que tiene y algún "pensamiento gemelo". Creemos que esto es suficiente para al menos sospechar que aceptaría la premisa (1) del argumento de Boghossian. La cosa es un poco más complicada con Schiffer. Primero, mencionemos que al exponer el argumento de Boghossian hemos dejado fuera algunos aspectos, ya que no nos parecían de suficiente interés para las cuestiones que queremos tratar en este trabajo. Boghossian (1992a) defiende que los ejemplos de transición lenta demuestran que el externismo semántico es incompatible con la transparencia del contenido y que por lo tanto hay términos y enunciados tipo en *mentalés* que son ambiguos. Por ejemplo, según Boghossian, si el externismo semántico fuera verdadero, el término 'Pavarotti' en el vocabulario mentalés de Peter no tendría una referencia o un contenido *permanentes*; el término sería ambiguo, diferentes ejemplares del término tendrían referencias y contenidos distintos. Por ello, si fuéramos más fieles a la letra de Boghossian (1992a), tendríamos que transcribir su argumento del siguiente modo:

(1) Si el externismo semántico es verdadero, entonces el contenido no es transparente.
(1.5) Si el contenido no es transparente, entonces hay términos y enunciados tipo en mentalés que son ambiguos.

años atrás se encontró con Pavarotti en el Lago Taupo, se basa en un recuerdo proveniente de sus experiencias en la Tierra y, por eso, está en parte constituida por el concepto PAVAROTTI. Al tener Peter la intención de llevar a cabo una inferencia, la segunda premisa hereda ese mismo concepto; así, la segunda premisa del argumento de Peter es que anoche escuchó cantar a Pavarotti (no a bi-Pavarotti). Y de estas dos premisas sí se sigue la conclusión de que una vez se encontró en el Lago Taupo con el tenor que escuchó cantar anoche—el argumento es válido, aunque la segunda premisa y la conclusión son falsas. Peter puede saber *a priori* qué se sigue de qué.

Schiffer (1992) además menciona lo que se supone es un argumento a favor de su propuesta. Peter estaría dispuesto a exponer las premisas de su argumento del siguiente modo "Una vez me encontré en el Lago Taupo a Pavarotti, a quien escuché cantar anoche", donde es explícito que en la segunda premisa se refiere al sujeto (a Pavarotti) anafóricamente, preservando la referencia de la primera premisa. El que Peter esté dispuesto a presentar el argumento en cualquiera de los dos modos demuestra, se supone, que en la segunda premisa hay una referencia anafórica en ambos casos (en el segundo caso la anáfora es implícita) y que, por lo tanto, a lo largo de toda la inferencia se hace referencia al mismo individuo y se emplean los mismos conceptos.

Por eso, es falso que si el externismo es verdadero entonces nuestras adscripciones *de dicto* no pueden racionalizar las creencias y conductas de los sujetos de las adscripciones. Primero, parece evidente que en el modelo de Schiffer y Burge no hay peligro de que Peter esté condenado a ser irracional—para cualquier inferencia deductiva que acepte Peter, las premisas y conclusiones subsiguientes heredan los

---

(2') Si hay términos y enunciados tipo en mentalés que son ambiguos, entonces uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias.

(C_1) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán explicar que es racional.

(C_2) Si uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias, entonces las adscripciones *de dicto* de esos pensamientos no podrán racionalizar su conducta.

Schiffer niega que los ejemplos de Boghossian demuestren que haya términos y enunciados tipo en mentalés que sean ambiguos. Aunque acepta que el término 'Pavarotti' es ambiguo en el vocabulario de Peter (puede referirse tanto a Pavarotti como a bi-Pavarotti), no cree que haya motivos para tener que aceptar que también haya términos ambiguos en mentalés; 'Pavarotti' en boca de Peter tiene dos términos distintos como traducción al mentalés. Por lo tanto, el consecuente de (1.5) es falso. Ahora, ¿rechazaría Schiffer (1) o (1.5)? ¿Qué opina sobre la (in)compatibilidad del externismo semántico y la transparencia del contenido? No es explícito al respecto, y por eso tenemos dudas. Aún así, y dado que cree que después de la transición Peter tiene tanto el concepto PAVAROTTI como el concepto BI-PAVAROTTI, nos inclinamos a pensar que aceptaría la premisa (1) y negaría la premisa (1.5).

conceptos de la primera premisa y, por lo tanto, no hay peligro de que el externismo sea motivo de que Peter acepte como válida alguna inferencia que no lo es. La mera reflexión *a priori* le basta a Peter para saber qué argumento es válido y cuál no.

Y tampoco tiene problemas este modelo con mostrar cómo pueden nuestras adscripciones explicar la conducta del sujeto. Sobre la bufanda que manda Peter a bi-Pavarotti, Burge y Schiffer podrían defender que Peter ha seguido un razonamiento que expresaría profiriendo la siguiente alocución: 'A Pavarotti le gustan las bufandas de colores; quiero hacerle un bonito regalo a Pavarotti; he de comprar una bufanda de colores y mandarla al domicilio de Pavarotti'. Dado que la primera premisa contiene el concepto PAVAROTTI, también lo contienen las demás premisas y la conclusión, el concepto BI-PAVAROTTI no aparece en el razonamiento que produce la conducta. Esto es, lo que explica la conducta de Peter son: su deseo de hacerle un regalo a Pavarotti y sus creencias de que a Pavarotti le gustan las bufandas de colores y de que Pavarotti vive en tal y cual lugar[212]. Parece que una de las creencias que tiene Peter, su creencia de que bi-Pavarotti vive en tal y cual lugar, es relevante para explicar su conducta sólo en cuanto en parte origina la creencia de que Pavarotti vive en ese lugar.

Schiffer (1992) dice que la intención del sujeto de seguir un argumento asegura que las premisas y la conclusión comparten conceptos; para acabar con la descripción de la propuesta que estamos presentando, mencionemos que Burge (1998) aduce que es la memoria preservativa la que posibilita este hecho. Ya hemos explicado en la segunda parte que, según Burge, la función principal de la memoria preservativa es preservar el punto de vista del sujeto a lo largo del tiempo; una de las tareas de la memoria preservativa sería *enlazar* entre sí los distintos pasos de un razonamiento, asegurándole al sujeto un mismo punto de vista a lo largo de él. Así, opina que "la memoria

[212] Sólo mencionemos que la respuesta de Schiffer y Burge tiene una consecuencia curiosa: al adquirir el concepto BI-PAVAROTTI la víctima de la transición lenta multiplica por dos las creencias de 'Pavarotti'. Supongamos que la primera creencia que se forma Peter al adquirir el concepto BI-PAVAROTTI es que bi-Pavarotti es un *tifossi* de la Juventus. Para cualquier creencia *p* de Pavarotti que tenga Peter, puede formar una inferencia tal que la primera premisa del argumento sea que bi-Pavarotti es un *tifossi* de la Juventus y la segunda una creencia *q*, tal que *q* es la proposición que resulta al cambiar el concepto PAVAROTTI que en parte constituye *p* por el concepto BI-PAVAROTTI. Así, para toda creencia *p* de Pavarotti que tiene Peter, también tiene una creencia *q* que se forma al cambiar el concepto PAVAROTTI de *p* por el concepto BI-PAVAROTTI (y viceversa: para toda creencia *q* de bi-Pavarotti que tiene Peter también tiene una creencia *p* de Pavarotti). Alguien podría pensar que este resultado es problemático o poco deseable (Tye (1998) lo piensa), no así nosotros. Ya que también la alternativa que defenderemos tiene esta misma consecuencia, será entonces cuando nos centremos en ella (sección 3.3.).

preservativa (...) es fundamental a la coherencia de la actividad racional"[213], y que "la memoria preservativa es necesaria para cualquier razonamiento que tenga lugar en el tiempo, por lo tanto para cualquier razonamiento"[214]. De hecho, tanto la memoria sustancial como la preservativa toman parte en los razonamientos:

> Para determinar qué está pensando Peter en la segunda premisa, uno debe recordar que tanto la memoria sustantiva de la experiencia de ayer como la memoria preservativa del uso del concepto en la primera premisa están operando en el pensamiento de Peter. Seguir el hilo de un argumento válido en el pensamiento es conectar premisas, manteniéndolas juntas de un modo que apoye la conclusión. Y la memoria preservativa – incluso en argumentos cortos que idealizamos como ocurriendo en un presente ilusorio – es esencial para este cometido. En cuanto tenemos la intuición de que Peter no está haciendo un error en su razonamiento, es natural, y en la mayoría de los casos creo que correcto, interpretar a Peter como manteniendo constante a lo largo del argumento, mediante la memoria preservativa, el concepto usado en la primera premisa en su pensar de la segunda premisa. El rol del concepto PAVAROTTI en el razonamiento es primario en su pensamiento, y la memoria preservativa toma la aparición del concepto en la primera premisa como base para su reempleo en la segunda premisa.[215]

Esto es, si la segunda premisa del argumento que piensa Peter es el pensamiento de que ayer escuchó cantar a Pavarotti (no a bi-Pavarotti), eso es así porque tanto la memoria sustancial como la preservativa toman parte en la tarea de "activar" ese pensamiento. La memoria sustancial es necesaria para que Peter piense un pensamiento que expresaría profiriendo el enunciado 'Anoche escuché cantar a Pavarotti', y la memoria preservativa para en parte determinar el contenido de ese pensamiento (como un contenido en parte constituido por PAVAROTTI) y así posibilitar que Peter lleve a cabo un razonamiento.

Brevemente, Schiffer (1992) y Burge (1998) niegan que el externista esté comprometido a mantener que la víctima de una transición lenta no podría conocer *a priori* las propiedades lógicas de sus pensamientos, ni a que aceptaría como válidos argumentos

[213] Preservative memory (...) [is] fundamental to the coherence of rational activity. (Burge (1998), p. 358)
[214] Preservative memory is necessary to any reasoning that takes place over time, hence any reasoning. (Burge (1998), p. 363)
[215] In determining what [Peter] is thinking in the second premise, one must remember that both the substantive memory of yesterday's experience, and the preservative memory of the use of the concept in the first premise, are operating in [Peter's] thinking. What it is to carry out valid arguments in thought is to connect premises, holding them together in a way that supports the conclusion. And preservative memory – even in short arguments that we idealize as occurring in a specious present – is essential to this enterprise. Insofar as we think intuitively that [Peter] is not making a mistake in reasoning, it is natural, and in most cases I think correct, to take [him] to be holding constant, through preservative memory within the argument, the concept used in the first premise in [his] thinking the second premise. The role of the concept PAVAROTTI in the reasoning is primary in [his] thinking, and preservative memory takes the occurrence of the concept in the first premise as a basis for its reuse in the second premise (Burge (1998), p. 367)

que no lo son. Eso es así, defienden, porque la intención argumentativa del sujeto causa que las ulteriores premisas y la conclusión hereden los conceptos que constituyen la primera premisa. Es falso que aquellos que sufren transiciones lentas estén condenados a ser irracionales, así como que nuestras adscripciones *de dicto* de actitudes proposicionales no puedan racionalizar tanto las creencias como la conducta del sujeto.

2.2.1 Problemas.

La estrategia Schiffer-Burge es, seguramente, la más conocida entre las respuestas ofrecidas al argumento de Boghossian y, por eso, también es la que ha recibido más atención. Son varios los autores que han mostrado dudas; nos limitaremos a poco más que mencionar y citar esas dudas y críticas. Creemos que muchas de estas críticas minan el carácter intuitivo que podría tener la propuesta y, además, nos parece que Collins (2008) presenta un argumento que, tenemos la impresión, difícilmente pueden responder Schiffer y Burge.

Es evidente que la propuesta tiene una consecuencia que, aunque no fatal, sí la provee de cierto matiz anti-intuitivo. Supongamos que en la mañana siguiente del recital, Peter piensa para sus adentros un razonamiento que expresaría profiriendo estas palabras: "Una vez me encontré con Pavarotti en el Lago Taupo; anoche escuché cantar a Pavarotti; por lo tanto una vez me encontré en el Lago Taupo al tenor que escuché cantar anoche". La primera premisa del argumento está en parte constituida por el concepto PAVAROTTI, y la segunda premisa hereda ese mismo concepto. Pero ahora supongamos que, cinco minutos más tarde, Peter piensa un razonamiento que expresaría profiriendo estas palabras: "Anoche escuché cantar a Pavarotti; una vez me encontré con Pavarotti en el Lago Taupo; por lo tanto una vez me encontré en el Lago Taupo al tenor que escuché cantar anoche". En este caso la primera premisa del argumento es evidentemente sobre bi-Pavarotti y, por lo tanto, las dos premisas del argumento estarán en parte constituidas por el concepto BI-PAVAROTTI. Esto es,

> La posición de Burge tiene la anti-intuitiva consecuencia de que el orden en el que un sujeto razona podría afectar el contenido de los pensamientos sobre los que se

> razona. (…) Burge está obligado a negar la intuitivamente plausible idea de que qué pensamientos tiene un sujeto es independiente del orden de su razonamiento.[216]

En el ejemplo que hemos esbozado, Peter estaría expresando dos razonamientos distintos, el primero de ellos sobre Pavarotti, el segundo sobre bi-Pavarotti—resulta extraño. Quisiéramos creer que, en principio, el orden de las premisas no debería alterar la identidad del razonamiento. Es más, tómese el primer razonamiento de Peter, el razonamiento sobre Pavarotti: parece que en la propuesta de Burge y de Schiffer Peter sería incapaz de pensar ese mismo razonamiento pero cambiando de orden las dos premisas. Esto es raro.

En una respuesta al artículo de Schiffer (1992), Boghossian (1992b) insinúa que la estrategia esbozada es, al fin y al cabo, una no-respuesta a su argumento:

> Ahora, imaginemos un caso—sin lugar a dudas, concebible—en el que Peter pone juntas dos de tales creencias mnemónicas, una proveniente de una experiencia en la Tierra y la otra de una experiencia en la Tierra Gemela, para llegar a una conclusión de identidad a la cual no había llegado antes. No llego a ver qué podría evitar esto.[217]

Parece que Boghossian está sugiriendo lo siguiente: incluso si aceptáramos que la intención de Peter de llevar a cabo un razonamiento y, por eso, de ceñirse a un uso no ambiguo de 'Pavarotti', evita que las dos premisas de su argumento estén compuestas por conceptos diferentes, no se sigue que en algún momento Peter no tenga la intención de "escoger" dos de sus creencias, asegurando así la identidad entre cada premisa, por un lado, y un recuerdo que tenía con anterioridad, por el otro. Si fuera posible que Peter pudiera comparar dos de sus creencias de este modo y llevar así a cabo una inferencia (Boghossian "no llega a ver qué podría evitar esto"), entonces seguiría siendo posible que Peter aceptara como válidos argumentos que de hecho no lo son, y que la reflexión *a priori* no fuera suficiente para hacerle ver su error.

---

[216] Burge's position has the counterintuitive consequence that the order in which a subject reasons may affect the content of the thoughts reasoned about. (...) Burge is forced to deny the intuitively plausible view that what thoughts a subject entertains is independent of the order of her reasoning. (Brown (2004), pp. 177-178)

[217] Now, just imagine a case—surely conceivable—in which Peter puts together two such memory beliefs, one stemming from an Earthly experience and the other from a Twearthly experience, in order to draw an identity conclusion he had not arrived at previously. I fail to see what could conceivably preclude this. (Boghossian (1992b), p. 42)

Burge (1998) contempla esta posibilidad, pero señala que no se seguiría que el sujeto es irracional, ya que una presuposición sobre la identidad entre Pavarotti y bi-Pavarotti salvaría la racionalidad del sujeto:

> En cuanto las intenciones del razonador en el razonamiento no sean dominantes en requerir "anafóricamente" que el mismo concepto sea usado a lo largo del razonamiento, y en cuanto pensemos que hay un salto entre las premisas que el razonador no ha hecho explícito, parecería obvio que el razonador tácita y erróneamente presupone que los conceptos se aplican a los mismos objetos. (...) Así, para capturar por completo el estado cognitivo del razonador en un caso donde el razonador presupone (erróneamente) que los conceptos se aplican a los mismos objetos, uno debería suplementar al razonador la presuposición errónea de que Pavarotti es bi-Pavarotti. De nuevo, no hay error en el razonamiento, sólo un error de presuposición.[218]

Lo que encontramos extraño aquí es que Burge no generalice esta respuesta. Si la presuposición de Peter de que Pavarotti es bi-Pavarotti hace que el argumento que acepta sea válido, no vemos por qué no acude Burge a estas presuposiciones en todas las inferencias problemáticas dejando de lado la estrategia basada en la anáfora y la herencia de conceptos—no vemos que haya motivos independientes para adoptar la posición en favor de la anáfora implícita.

Hemos mencionado antes que Schiffer argüía que la predisposición de Peter a cambiar la forma de su argumento, en un modo tal que se hace una referencia anafórica explícita, mostraba que en la forma original del argumento sí había una referencia anafórica (más o menos) implícita.

> Bueno, el hecho de que Peter aceptaría con agrado reformular su inferencia sustituyendo 'Pavarotti' por 'él' no muestra que nada de lo que haya hecho garantiza que todas las apariciones del término 'Pavarotti' co-refieren. Simplemente refleja el hecho de que cree que lo hacen – y *por supuesto* lo cree, o si no, no haría la inferencia.[219]

---

[218] Insofar as the reasoner's intentions in reasoning are not dominant in requiring "anaphorically" that the same concept be used through the reasoning, and insofar as we think that there is a gap between the premises that the reasoner has not made explicit, it would seem obvious that the reasoner tacitly and mistakenly presupposes that the concepts apply to the same objects. (...) So to fully capture the reasoner's cognitive state in a case where the reasoner does presuppose (mistakenly) that the concepts apply to the same objects, one would have to supply for the reasoner the mistaken presupposition that [Pavarotti is bi-Pavarotti]. Again, there is no mistake in reasoning, only a mistake in presupposition. (Burge (1998), p. 368)

[219] Well, the fact that Peter would happily enough rephrase his inference substituting ['he'] for ['Pavarotti'] doesn't show that anything he's done has guaranteed that the occurrences of the word ['Pavarotti'] all co-refer. It just reflects the fact that he thinks they all do – and *of course* he thinks that, or else he wouldn't make the inference. (Collins (2008), p. 562)

Una vez aceptamos que el contenido no es transparente, abrimos la puerta a que uno contemple dos razonamientos distintos, que uno de ellos no sea válido y haga referencia a Pavarotti y a bi-Pavarotti, que el otro sí sea válido y haga referencia sólo a Pavarotti mediante anáfora, y que uno erróneamente crea que se trata del mismo argumento. Mientras este escenario sea posible, la disposición de Peter a reformular su argumento haciendo una referencia anafórica explícita no será motivo para pensar que en el argumento original hay una referencia anafórica implícita.

Pasemos, ya para terminar, a la última crítica que queríamos mencionar. La cosa es que, sea estipulando memorias preservativas y no preservativas, o sea mediante algún otro mecanismo, Schiffer y Burge tienen que explicar cómo llega Peter a la creencia de que anoche escuchó cantar a Pavarotti, habiendo partido de su creencia de que anoche escuchó cantar a bi-Pavarotti. Según Collins (2008), la única explicación posible es que Peter llega a esa creencia mediante *inferencia*:

> Si los contenidos de mis creencias son en primera instancia independientes los unos de los otros, pero cuando se ponen juntas en una inferencia algo les *sucede*, tal que se garantiza que sus contenidos se relacionan en el modo relevante, parece que esta cosa que les sucede es una *inferencia* (incluso si es inconsciente). Si, de mis creencias originales de *que nadé en agua cuando era crío* y *que ahora hay bi-agua condensándose en mi vaso*, de algún modo termino con una inferencia en la que las premisas están ligadas mediante anáfora, como *el agua es tal que nadé en ella cuando era crío; y ahora se está condensando en mi vaso*, esta transición sería el resultado de una inferencia (no-válida).[220]

Tanto Burge como Schiffer defienden que Peter recuerda que anoche escuchó cantar a bi-Pavarotti. Parece evidente, además, que ese recuerdo, esa creencia, está en el origen del pensamiento que constituye la segunda premisa del argumento de Peter, que anoche escuchó cantar a Pavarotti. Burge y Schiffer tienen que explicar cómo *transita* Peter de una creencia a la otra. Según Collins, y creemos que en esto tiene razón, la única explicación posible es que Peter *infiere* de su recuerdo, de su evidencia mnemónica, que anoche escuchó cantar a Pavarotti. Pero uno no debería inferir de su recuerdo de que anoche escuchó cantar a bi-Pavarotti que anoche escuchó cantar a Pavarotti, porque tal

[220] If the contents of my beliefs are independent of each other in the first place, but when they are put together in an inference something *happens* to them, so that their contents are guaranteed to match in the relevant ways, this thing that happens to them seems to be *inference* (even if it is unconscious). If, from my original beliefs *that I swam in water as a child* and *that now twater is condensing on my glass*, somehow I end up with an inference where the premises are anaphorically linked, like *water is such that I swam in it as a child; and now it is condensing on my glass*, this transition would be the result of an (invalid) inference. (Collins (2008), p. 562)

inferencia no es válida. Schiffer y Burge tienen que explicar cómo *transita* Peter de su recuerdo a la segunda premisa del argumento, y acudir a relaciones anafóricas implícitas no les servirá de nada. No han solucionado el problema que señalaba Boghossian[221].

Por eso, creemos que es preferible buscar en otra dirección.

[221] Podría responderse, teniendo en mente la cita de Burge (1998) de más arriba, que en este caso es una presuposición de identidad la que hace el trabajo. Como hay una presuposición de identidad (que, probablemente, tiene *status* de premisa), la inferencia de Peter de que anoche escuchó a bi-Pavarotti y, por lo tanto, anoche escuchó a Pavarotti, es válida. Creemos que esta respuesta es buena, así lo defenderemos en el tercer capítulo. El problema es que, si esto es así, entonces la anáfora implícita, por sí sola, no esquiva el problema que identificaba Boghossian. Lo que hace que el razonamiento de Peter sea válido es el carácter de presuposición o premisa que adquiere en algún momento su creencia de que bi-Pavarotti es Pavarotti.

3. CREENCIAS DE IDENTIDAD E INFERENCIA

En este capítulo describiremos la que creemos es la respuesta más adecuada al argumento de Boghossian. Diremos que el ejemplo de Peter no presenta ningún problema importante para el externista, lo único que sucede es que la interpretación que hace Boghossian del ejemplo es errónea—el externista tiene a mano una descripción de Peter que explica perfectamente por qué es racional y por qué actúa como lo hace.

Hay algunos ejemplos que guardan semejanzas importantes con el de Peter, aunque son más mundanos y comunes (no contienen ningún elemento extravagante del estilo de una transición lenta)—en los siguientes párrafos presentaremos dos. Creemos que el de Peter no es más que un ejemplo (algo retorcido) de este tipo de situaciones y que, si los casos más comunes no suponen un gran problema para nuestras prácticas habituales de adscripción de actitudes, tampoco lo hará el ejemplo de Peter. Por eso, nos basaremos en este tipo de ejemplos para defender que el externista puede explicar sin mayores problemas la situación de Peter si acude a alguna premisa oculta de no-identidad en el razonamiento que sigue éste.

3.1. TURISTAS, FILÓSOFOS Y CANTANTES

3.1.1. Pepe: filósofo y cantante.

Gorka, Miguel y Joan, becarios escribiendo su tesis doctoral, son miembros del departamento de Filosofía. Después de un largo día de trabajo, acuden a un bar a tomar algo, donde Leire, compañera de piso de Gorka, se incorpora al grupo. Gorka, Miguel y Joan no dejan de hablar sobre sus estudios, inquietudes filosóficas y compañeros de departamento, y claro, Leire les escucha aburrida. Comienzan a hablar de Pepe, compañero del departamento, quien ha escrito un artículo sobre la función semántica del punto y coma (interesantísimo, según opina Miguel). Leire escucha sin decir nada, recapacitando por qué querría Pepe (a quien no conoce) escribir un artículo sobre un tema semejante. A lo largo de la tarde tratan varios temas, Miguel y Joan discuten sobre la posibilidad de una teoría naturalizada de la mente, Joan les cuenta anécdotas de su estancia en una universidad de Montana, y Gorka les confiesa que no entendió nada del seminario del día anterior. Al final de la tarde, Joan les pregunta a los otros dos si han visto alguna vez tocar en directo a Pepe, compañero del departamento, quien canta en un grupo llamado *Caetano*. "Vaya," piensa para sus adentros Leire, "Pepe, compañero de departamento de Gorka, es el cantante de Caetano."

Recapacita, y llega a la conclusión de que el cantante de Caetano escribe artículos sobre la semántica del punto y coma.

La conclusión a la que llega Leire es falsa, el cantante de Caetano no escribe artículos sobre la semántica del punto y coma. Hay dos personas que se llaman 'Pepe' en el departamento de Gorka: uno de ellos escribe artículos sobre la semántica del punto y coma, el otro canta en un grupo de pop-rock.

3.1.2. John: turista y despistado.

John es originario de Nebraska, y gran aficionado al fútbol (sigue con atención todas las ligas europeas). Todos los lunes al llegar al trabajo repasa los resultados de los partidos disputados, y no pierde detalle de lo que han hecho equipos como el Barça, el Real

Madrid, el Manchester United o el Inter de Milán. Pero, cosa extravagante, John siente cierta debilidad por la Real Sociedad. Conoce los integrantes de la plantilla y las alineaciones, lee las crónicas de los partidos que juega la Real, y su mayor deseo es que el Athletic de Bilbao descienda a segunda división. John sueña con ir a San Sebastián y ver un partido de la Real (por lo demás, prácticamente lo único que sabe de San Sebastián es que es la ciudad donde juega la Real Sociedad).

Un día, John viaja a Europa por motivos de trabajo; acude a un congreso que se celebra en Pamplona. Descubre que San Sebastián está cerca de Pamplona, lo cual le llena de emoción: no trabaja el domingo por la tarde, quizás por fin pueda ver en vivo cómo juega la Real Sociedad. Decide que alquilará un coche e irá a San Sebastián si (y sólo si) la Real juega allí ese domingo, y compra un periódico para mirar dónde juega la Real ese fin de semana. Apenado, descubre que la Real juega en Donostia. No alquila ningún coche, se pasa el domingo por la tarde paseando por las calles de Pamplona, y el lunes toma el avión de regreso a Nebraska, triste y abatido.

John no ha ido a San Sebastián porque ha llegado a la conclusión de que ese domingo la Real no juega allí. La conclusión es falsa: Donostia es San Sebastián, y la Real jugaba en San Sebastián (Donostia) ese domingo.

3.1.3. Comparaciones.

Los ejemplos de Leire y John son bastante comunes, todos nos hemos encontrado en alguna situación parecida en algún momento. Además, opinamos que los dos guardan similitudes importantes con el ejemplo de Peter, hasta el punto que podemos tratar la situación de Peter como normalmente tratamos ejemplos como el de John.

Primero, los tres llegan a una conclusión falsa partiendo de (algunas) premisas verdaderas. El cantante de Caetano no escribe artículos sobre la semántica del punto y coma, pero sí es verdad que Gorka ha dicho que Pepe escribe artículos de ese tipo y que también ha dicho que Pepe es el cantante de Caetano; es falso que la Real no juegue en San Sebastián, pero cierto que juega en Donostia ese mismo día; Peter no vio en el Lago Taupo al tenor que escuchó, pero sí vio a Pavarotti en el Lago Taupo, y escuchó anoche a bi-Pavarotti.

Además, son esas creencias verdaderas que hemos mencionado las que (en parte) llevan a sus sujetos a adoptar la creencia falsa en cuestión; esto es así porque hay una inferencia de por medio. Leire no ha obtenido evidencia que inmediatamente justifique su creencia de que el cantante de Caetano escribe artículos de semántica (Gorka no ha proferido el enunciado 'El cantante de Caetano escribe artículos sobre semántica'); ha llegado a esa creencia infiriéndolo de las otras creencias que ha adquirido (justificadamente) con anterioridad. John tampoco ha obtenido evidencia que inmediatamente justifique que la Real no juega en San Sebastián (en el periódico no ponía 'La Real no juega en San Sebastián este domingo'), lo ha inferido (en parte) de su creencia (justificada) de que la Real juega en Donostia ese día.

A pesar de ello, queremos decir que los tres, Leire, John y Peter, se están comportando de manera racional. Y queremos decir que son racionales, entre otras cosas, porque creemos que tienen justificación para adoptar la creencia falsa a la que se adhieren (Leire "tiene motivos" para pensar que el cantante de Caetano escribe artículos sobre el punto y coma, John para creer que la Real no juega en San Sebastián, Peter para afirmar que el tenor que vio en el Lago Taupo es el tenor que escuchó anoche).

Esto es, los tres llegan a creencias falsas mediante inferencias que contienen creencias verdaderas y justificadas como (algunas de las) premisas, y queremos decir que los tres son racionales (si no, no estarían justificados en adoptar la creencia falsa en cuestión)—y si son racionales, de acuerdo con los criterios de Boghossian, la inferencia en cuestión que aceptan habrá de ser válida. Llegados a este punto, es necesario ofrecer una explicación de cómo es que alguien como Peter o John pueda tener una creencia falsa pero justificada producto de una inferencia que contiene premisas verdaderas.

Un último apunte sobre las similitudes entre los tres casos: los tres personajes tienen que hacer el mismo descubrimiento para corregir su error; los tres tienen que descubrir la falsedad de una creencia de (no-)identidad que tienen—Leire que Pepe (el articulista) no es Pepe (el cantante), John que Donostia es San Sebastián y Peter que Pavarotti no es bi-Pavarotti. Esto nos da una pista importante (creemos) sobre cuál debe ser la explicación correcta de las tres situaciones.

3.2. ESTRATEGIA DE LA PREMISA OCULTA

Comencemos por centrarnos en el caso de John; éste es un caso menos extravagante que el de Peter, y afirmamos que lo que digamos para explicar la racionalidad de John nos servirá también en el caso del otro. Uno podría afirmar que John llega a la conclusión de que la Real no juega en San Sebastián porque razona según la siguiente inferencia—a esta interpretación del ejemplo de John la llamaremos *la interpretación de Boghossian*, por ser análoga a las opiniones de Boghossian sobre Peter:

i. La Real juega en Donostia este domingo.
ii. Ningún equipo juega en dos ciudades distintas el mismo día.
iii. Por lo tanto, la Real no juega en San Sebastián este domingo.

Basándose en este razonamiento, uno podría explicar la conducta de John acudiendo a su deseo de acudir a San Sebastián si y sólo si la Real juega allí ese domingo y su creencia de que la Real juega en Donostia ese domingo.

Ahora, sucede que el razonamiento que hemos descrito no es válido (de hecho, las premisas son verdaderas pero la conclusión falsa), y que la explicación de la conducta de John que hemos dado tampoco es buena (compárese: Jon, natural de Pamplona, iría a San Sebastián si tuviera exactamente la misma creencia y el mismo deseo de John). Por eso, *la interpretación de Boghossian* de la situación de John no es buena; queremos decir que es racional, pero lo describe como aceptando una inferencia que no es válida.

Nadie propondría ni aceptaría *la interpretación de Boghossian* en el caso de John; la posición de éste es fácil de racionalizar, nunca se ha temido que ejemplos como éste pongan ningún problema para las adscripciones *de dicto* de nadie. Como sabemos, podemos explicar fácilmente la racionalidad y la conducta de John si acudimos a su creencia de que Donostia no es San Sebastián. Si a las premisas del razonamiento les añadiéramos la premisa de que Donostia no es San Sebastián, la conclusión se seguiría—ésta sería falsa, pero el razonamiento válido (la premisa de que Donostia no

es San Sebastián también es falsa). Si al deseo de John de ir a San Sebastián si y sólo si la Real juega allí, y a su creencia de que la Real juega en Donostia, le añadiéramos su creencia de que Donostia no es San Sebastián, esto explicaría perfectamente por qué actúa John del modo en que lo hace (véase: esa creencia que no tiene Jon explica por qué actúan Jon y John de diferente modo).

Si esta segunda interpretación (la común) es acertada, entonces nuestras adscripciones *de dicto* explican sin mayores problemas que John es racional, así como su conducta.

Quizás, el defensor de *la interpretación de Boghossian* podría intentar aferrarse a su explicación alegando que, si le preguntáramos a John, éste respondería que adquiere la creencia de que la Real no juega en San Sebastián partiendo sólo de sus creencias de que la Real juega en Donostia y que ningún equipo juega en dos ciudades distintas el mismo día (podemos estipular que John respondería así si se le preguntara). Y esto, se supone, sería evidencia a favor de que la inferencia concreta que de hecho sigue John no contiene una premisa de no-identidad.

Este movimiento suena un tanto desesperado, descansa sobre una interpretación errónea de qué determina cuáles son las premisas que constituyen las inferencias que considera uno. Constantemente razonamos mediante inferencia, muchas veces (la mayoría) de forma inconsciente. Cuando conscientemente razonamos mediante inferencia, es común que no pensemos conscientemente todas y cada una de las premisas que de hecho constituyen el razonamiento que seguimos—esperamos que esto sea un punto compartido por todos, no entraremos a justificarlo. Por eso, uno puede tener creencias falsas sobre cuáles son las premisas que en parte constituyen un razonamiento concreto que sigue de manera consciente; no se sigue que la opinión de uno sobre qué premisas toman parte en las inferencias que acepta sea infalible.

S llega a la creencia de que *q* mediante inferencia (parcial o totalmente inconsciente o no), y nosotros queremos explicar que S es racional. Para ello tenemos que identificar cuál es la inferencia deductiva válida que de hecho sigue S. Y la inferencia que de hecho sigue S tendrá como premisa toda creencia *p* tal que: *p* es una creencia de S, y *p* ha motivado en parte que S adopte la creencia de que *q* (en el sentido de que, de no ser *p* una creencia de S, éste no habría adoptado la creencia de que *q*). El que S crea o no que

una de sus creencias *p* de hecho constituye una de las premisas de la inferencia, no es relevante—perfectamente puede tener creencias falsas al respecto.

Por eso, tenemos a mano una buena explicación de la racionalidad de John. Su creencia de que Donostia no es San Sebastián constituye una premisa en la inferencia que de hecho sigue porque, primero, es una creencia que tiene John y, segundo, si John no tuviera esa creencia no habría llegado a la conclusión de que la Real no juega en San Sebastián. No hay *puzzle* en el caso de John, lo que había era una mala explicación (*la interpretación de Boghossian*)—exactamente lo mismo proponemos nosotros en el caso de Peter.

Es fácil dar una respuesta análoga al ejemplo de Peter. Así, *la estrategia de la premisa oculta* afirma que el razonamiento que sigue Peter contiene la siguiente premisa (0), y que, por eso, Peter es racional:

(0) Pavarotti es bi-Pavarotti.
(1) Una vez me encontré con Pavarotti en el Lago Taupo.
(2) Ayer escuché cantar a bi-Pavarotti.
(3) Por lo tanto, una vez me encontré en el Lago Taupo al tenor que escuché anoche.

El argumento, así expuesto, sí es válido, (3) se sigue de (0), (1) y (2)[222]. Si la inferencia que considera Peter de hecho contiene la premisa (0), entonces no parece que Peter acepte como válida ninguna inferencia que de hecho no lo es.

Además, resulta fácil explicar la conducta de Peter si tenemos en cuenta esta creencia de que Pavarotti es bi-Pavarotti. Peter manda una bufanda de colores a bi-Pavarotti porque quiere hacerle un bonito regalo, y cree que a Pavarotti le gustan las bufandas de colores, y que Pavarotti es bi-Pavarotti. Si ésta no es una buena explicación de la conducta de Peter, no llegamos a entender qué lo podría ser.

[222] Así lo atestigua su forma lógica:
(1) $a = b$
(2) Fa
(3) Gb
(4) $\exists x (Fx \wedge Gx)$

Y, al menos en principio, parece que Peter sí tiene esta creencia de que Pavarotti es bi-Pavarotti (sí al menos si nos adherimos a un modelo de cohabitación de conceptos). Al fin y al cabo, Peter estaría dispuesto a proferir enunciados como 'Cuando pienso en el tenor que vi en el Lago Taupo y pienso en el tenor que escuché anoche estoy pensando en la misma persona'. Creemos que esto es evidencia suficiente para adscribirle la creencia de identidad en cuestión.

Así, la estrategia de la premisa oculta viene a decir que la inferencia de Peter de hecho sí contiene la premisa (0), que una descripción de la inferencia que no contiene la premisa en cuestión es un *entimema*. La interpretación de Boghossian del ejemplo era equivocada porque no incorporaba una creencia que sí tiene Peter y que es crucial para explicar su racionalidad y su conducta. El externista tiene a mano una buena explicación de por qué es Peter racional, y de por qué actúa del modo en que lo hace—sus adscripciones *de dicto* cumplen con el cometido que se les había asignado. Al igual que con John, no hay *puzzle* en el ejemplo de Peter.

Hay una pregunta que ha de responder el defensor de la estrategia de la premisa oculta. Cuando un sujeto considera alguna inferencia en cuestión, ¿hay *siempre* premisas de (no-)identidad (ocultas) que toman parte en la inferencia? Si la respuesta es que no, ¿cuándo contiene una inferencia alguna premisa de ese tipo y cuándo no? Las respuestas a estas preguntas son: "No" y "Siempre que nos sea necesario". Expliquemos brevemente estas respuestas.

Supongamos que alguien llega a la creencia de que hay un objeto que es *F* y *G* partiendo de las premisas *Fa* y *Ga*; ¿contiene la inferencia una premisa de identidad *a=a*? No. La inferencia "*Fa*; *Ga*; Hay al menos un objeto que es *F* y *G*" es válida, una premisa de identidad sería superflua, y no necesitamos adscribir una creencia tal para explicar que alguien que sigue la inferencia es racional. Como no necesitamos acudir a esa creencia de identidad para explicar cómo es que el sujeto en cuestión es racional, como la inferencia que acepta ya es válida sin implementarla con una premisa de identidad, no hay motivos para suponer que la inferencia tiene una premisa de este tipo.

Así, es posible que S y S* sean dos sujetos internamente indistinguibles, que tengan creencias que expresarían con los mismo enunciados (tipo), que sigan inferencias que

darían a conocer exactamente con las mismas palabras, y que una de esas inferencias contenga una premisa de identidad en el caso de S, pero no así en el caso de S*. Pongamos un ejemplo.

Supongamos que Peter* ha tenido exactamente las mismas vivencias (internamente individuadas) que Peter, pero que guardan una diferencia: Peter* no ha sido víctima de una transición lenta. Para cada "vivencia de bi-Pavarotti" que ha tenido Peter, Peter* ha tenido una "vivencia de Pavarotti". Así, un día Peter* llega a la conclusión de que una vez se encontró en el Lago Taupo al tenor que escuchó anoche, porque una vez vio a Pavarotti en el Lago Taupo, y anoche escuchó cantar a Pavarotti. Las premisas y la conclusión son verdaderas, y el argumento, así expuesto, válido. ¿Contiene el razonamiento de Peter* una premisa (oculta) de identidad "Pavarotti es Pavarotti"? No. Y no la contiene, simplemente, porque no necesitamos de esa creencia para racionalizar sus creencias y explicar su conducta, una creencia de este tipo sería superflua en nuestra explicación.

Esto es, tanto Peter como Peter* tienen cada uno una creencia que expresarían profiriendo el enunciado 'Pavarotti es Pavarotti': Peter cree que Pavarotti es bi-Pavarotti y Peter* que Pavarotti es Pavarotti. Ambos son internamente indistinguibles y, por lo tanto, los dos siguen una inferencia que darían a conocer profiriendo las siguientes palabras: "Una vez me encontré con Pavarotti en el Lago Taupo; Anoche escuché cantar a Pavarotti; Por lo tanto, una vez me encontré en el Lago Taupo con el tenor que escuché cantar anoche". Ahora, los razonamientos que siguen y que darían a conocer mediante sus respectivas proferencias son distintas. Peter* sí cree que Pavarotti es Pavarotti, pero esta creencia no constituye una premisa en la inferencia que sigue, porque sería superflua, y porque no necesitamos de ella para explicar que Peter* es racional. En cambio, la creencia de Peter de que Pavarotti es bi-Pavarotti sí constituye una premisa en la inferencia que sigue, porque esta premisa no es superflua (la validez de la inferencia que sigue Peter depende de ella), y porque necesitamos de ella para explicar que Peter es racional.

Esto es, la estrategia de la premisa oculta tiene la siguiente consecuencia: cuando S tiene una creencia de identidad verdadera que expresaría profiriendo el enunciado '$a=a$', esa creencia no constituye premisa en las inferencias que sigue S; cuando S tiene

una creencia de identidad falsa que expresaría profiriendo el enunciado '*a*=*a*', esta creencia sí constituye una premisa en algunas de las inferencias que sigue. Hay factores externos (como la veracidad o falsedad de las creencias de identidad de uno) que son relevantes para determinar qué inferencias sigue uno, para determinar si una inferencia en cuestión que sigue tiene o no una premisa de identidad.

Algunos pueden pensar que esta es una idea difícil de asumir, nosotros creemos que es una consecuencia que el externista tendría que abrazar sin mayores complejos.

Por lo tanto, el externista tiene a mano adscripciones *de dicto* suficientes para explicar la racionalidad de Peter y su conducta. No hay *puzzle* en este ejemplo, ni problema para el externista, el supuesto embrollo se formó cuando Boghossian se olvido de una creencia de identidad de Peter a la hora de explicar su situación. Y ésta es una buena respuesta al argumento de Boghossian.

3.3. CRÍTICAS (Y RESPUESTAS)

Tye (1998) protesta que no es plausible que Peter tenga la creencia de que Pavarotti es bi-Pavarotti—contempla la posibilidad de que la víctima de una transición lenta tenga la creencia de que el agua es bi-agua, y afirma que no es una idea plausible.

Da dos razones para negar que Peter cree que el agua es bi-agua. Primero, dice que no es plausible que, tras muchos años en la Tierra Gemela, Peter crea que está rodeado de agua; lo que ciertamente cree Peter es, según él, que está rodeado de bi-agua. Ahora, si cree que está rodeado de bi-agua y, además, también cree que la bi-agua es agua, parece que Peter también tiene la creencia (disposicional) de que está rodeado de agua. Las creencias de 'agua' de Peter se multiplican por dos cuando adquiere el concepto BI-AGUA.

Bueno, no nos parece que quien apueste por la cohabitación de conceptos (como nosotros) tenga mayores problemas para morder esta bala, al fin y al cabo creemos que

es una consecuencia que se sigue naturalmente de la cohabitación. Puede que sea una idea poco intuitiva, pero menos intuitiva le parecerá seguramente al defensor de la cohabitación que Peter pierda todas las creencias y recuerdos de agua que tenía (lo cual es, como sabemos, una consecuencia del reemplazo conceptual). Probablemente, cuál de estas consecuencias le parezca a uno que es más fácil de asumir será uno de los motivos por los que apostará o bien por la cohabitación o bien por el reemplazo.

El otro motivo aducido por Tye es que Peter podría adentrarse en un experimento filosófico de Tierra Gemela. Podría imaginar que hay un planeta, muy similar al suyo, pero que contiene agua (a la que Peter referiría usando el término 'bi-agua'), una sustancia distinta a la bi-agua (a la cual él llama 'agua'). Preguntado sobre si la sustancia que contiene el planeta que él llama 'Tierra Gemela' es la sustancia que tiene alrededor, Peter diría que no, Peter tiene la disposición de proferir el enunciado 'La bi-agua no es agua'. Por lo tanto, concluye Tye, Peter no tendría la creencia de que el agua y la bi-agua son lo mismo.

Pero Tye construye el escenario de modo totalmente inadecuado. Cuando se estipulan escenarios de Tierra Gemela, se estipula que la sustancia acuosa que contiene la Tierra Gemela tiene una estructura química diferente al agua y, se supone, es esto lo que hace que esa sustancia acuosa (la bi-agua) no sea agua. Por lo tanto, parece, al adentrarse en el experimento filosófico, Peter debería estipular que lo que él llama 'bi-agua' es H_2O, o al menos que no es XYZ. Pero la cuestión es que si Peter sabe que la bi-agua es XYZ, entonces tiene conocimiento que diferencia el agua de la bi-agua, puede distinguir entre estas dos sustancias y, por eso, el suyo no sería un ejemplo de transición lenta (a las transiciones lentas les es esencial que la víctima confunda el agua con la bi-agua, o Pavarotti con bi-Pavarotti).[223]

No vemos por lo tanto razones suficientes para dejar de pensar que Peter sí cree que el agua es bi-agua, o que Pavarotti es bi-Pavarotti.

[223] Se protestará, con razón, que esto no es necesario, Peter puede estipular que lo que él llama 'bi-agua' es diferente a lo que tiene alrededor aun y cuando desconoce la estructura química de la bi-agua—basta con que piense que hay una sustancia muy parecida a lo que él llama 'agua', pero que de hecho no lo es. Pero esto tampoco sirve para conseguir los objetivos que persigue Tye; no hay motivos para pensar que la sustancia que estipula Peter es agua, podría ser "tri-agua" (ABC), o "tetra-agua" (BCD), o "penta-agua" (CDE), o... nada en el experimento de Peter, así expuesto, hace indicar que lo que contiene el planeta que él llama 'Tierra Gemela' es agua. Por lo tanto, no está diciendo que el agua no es bi-agua cuando profiere 'la bi-agua no es agua', porque del mismo modo podría estar diciendo que la bi-agua no es tri-agua, o que la bi-agua no es tetra-agua, o que...

Collins (2008) menciona un problema que cree que no es contundente, pero que nos gustaría traer a colación aquí. Parece que la estrategia que hemos defendido abre la puerta a que nadie (o casi nadie) sea irracional. Pongamos un ejemplo. Supongamos que Coraline acepta como válido el siguiente argumento:

1. Si el gato del jardín habla, entonces mi Otra Madre tiene botones en vez de ojos.
2. El gato de mi jardín no habla.
3. Por lo tanto, mi Otra madre no tiene botones en vez de ojos.

Claramente, el argumento no es válido, y Coraline se está comportando de manera irracional al aceptarlo. Pero parece que alguien podría afirmar, como hemos hecho nosotros con Peter y su creencia de identidad, que Coraline tiene la siguiente creencia 0, y que el argumento que acepta es válido porque contiene esa creencia como premisa:

0. Si, si el gato del jardín habla, entonces mi Otra Madre tiene botones en vez de ojos, y el gato de mi jardín no habla, entonces mi Otra madre no tiene botones en vez de ojos.

O puesto un poquito más claramente: Si [(si el gato del jardín habla, entonces mi Otra Madre tiene botones en vez de ojos) y el gato del jardín no habla], entonces mi Otra Madre no tiene botones en vez de ojos[224]. Por supuesto, si añadiéramos esta premisa al argumento, entonces éste sería válido, y de hecho es una creencia que sí tiene Coraline y que motiva que se adentre en la inferencia. Así, uno se puede preguntar qué es lo que nos hace pensar que, a diferencia de Peter, Coraline es irracional.

Es fácil dar una respuesta a esta crítica. Cabe preguntarse por qué adopta Coraline esa creencia, qué es lo que la justifica. Si la respuesta fuera que son las creencias concretas de Coraline sobre la naturaleza del gato y de su Otra Madre (y no sus creencias metalógicas sobre derivabilidad, inferencia o validez) las que hacen que adopte la creencia 0, entonces, dependiendo de las características concretas del ejemplo, de hecho podría defenderse que el argumento sí contiene la premisa 0 y que, por eso, es válido. Ahora, cabe la posibilidad de que Coraline tenga la creencia 0 porque cree falsamente

[224] Más fácil todavía: [(Gato habla→Otra Madre Botones) ∧ Gato no habla] → Otra Madre no Botones.

que de las premisas 1 y 2 de su argumento se sigue la premisa 3 sin ayuda de más premisas, porque tiene la creencia de que de la premisa P→Q y la premisa ¬P se sigue la conclusión ¬Q (esto es, Coraline cree P→Q; ¬P |- ¬Q). Y es justo esta creencia metalógica la que la hace irracional, a diferencia de Peter Coraline ni siquiera sabe qué estructuras argumentales son válidas. Por contra, las creencias de (no-)identidad que tienen John y Peter no son creencias metalógicas, tampoco creencias que se siguen de sus creencias metalógicas, son creencias que tienen sobre la identidad de los objetos sobre los cuales piensan y juzgan. Estas diferencias entre las creencias relevantes de Peter y Coraline explican claramente por qué el primero es racional pero la segunda no.

Brown (2004) y Faria (2009) objetan a la estrategia de la premisa oculta[225] que nos lleva a un regreso infinito[226]. Los dos asumen que "la teoría de la premisa de identidad ha de ser parte de una posición más general diciendo que un sujeto no puede simplemente descansar sobre la identidad de los conceptos en la inferencia"[227]. Brown acude así a un ejemplo parecido al de Peter* que hemos presentado antes, quien hace una inferencia análoga a la de Peter, pero no ha sido víctima de una transición lenta. Según la interpretación que hace Brown de la estrategia de la premisa oculta, la inferencia que sigue Peter* tiene una premisa implícita "Pavarotti es Pavarotti". Y esto es un problema, ya que el estratega de la premisa oculta cae en un regreso infinito:

> Pero incluso añadiendo esta premisa de identidad, el argumento todavía descansa sobre la identidad de los conceptos. En particular, descansa sobre el hecho de que el concepto expresado mediante el primer uso de ['Pavarotti' en 0: 'Pavarotti es Pavarotti'] es idéntico al concepto expresado mediante el uso de ['Pavarotti' en 1: 'Me encontré con Pavarotti en el Lago Taupo']; y, que el concepto expresado mediante el segundo uso de ['Pavarotti' en 0] es idéntico al concepto expresado mediante el uso de ['Pavarotti' en 2: 'Anoche escuché a Pavarotti']. Si uno no puede nunca descansar sobre la identidad de los conceptos en una inferencia, el argumento necesita suplementarse con ulteriores premisas de identidad. Pero hemos comenzado un regreso. Cualquier premisa de identidad adicional de la forma ['Pavarotti es Pavarotti'] descansa sobre la identidad entre el concepto expresado mediante el primer uso de ['Pavarotti'] en esta premisa y la expresada

[225] Tenemos dudas de que la respuesta que tienen en mente Brown y Faria y que quieren criticar sea exactamente la estrategia que hemos descrito nosotros. Si no lo fuera, entiéndase lo que sigue como una posible crítica a la estrategia de la premisa oculta que se basa en algunas opiniones de Brown (2004) y Faria (2009).

[226] Los dos citan a Campbell (1987) como proponiendo el germen de su crítica. Brown dice correctamente que las afirmaciones de Campbell se enmarcan en el contexto de otra discusión, que no propuso explícitamente la crítica que sigue; Faria sí acredita equivocadamente a Campbell como el primero en proponer la crítica que sigue contra la estrategia de la premisa oculta.

[227] ...the identity-premise view must be part of a more general view that a subject cannot just rely on the identity of concepts in inference. (Brown (2004), p. 180)

> mediante el uso de ['Pavarotti'] en alguna otra premisa. Si uno no puede nunca descansar sobre la identidad de los conceptos, la inferencia tendrá que contener una premisa de identidad ulterior. Pero la misma cuestión sale a superficie con toda premisa ulterior.[228]

Esta crítica no entiende bien cuál es el objetivo del argumento de Boghossian (ni, por lo tanto, la respuesta basada en la estrategia de la premisa oculta). Brown (2004) y Faria (2009) interpretan el argumento de Boghossian como diciendo que, si el externismo es verdadero, entonces alguien en la situación de Peter no puede saber *a priori* qué argumento es válido y cuál no, y entienden que la estrategia de la premisa oculta responde que Peter podría *asegurar* su racionalidad implementando su argumento con la premisa de identidad "Pavarotti es bi-Pavarotti" (algo que debería hacer con cada uno de sus argumentos). Pero ésta es una mala interpretación del argumento y la respuesta.

Ya dijimos en el primer capítulo que el argumento de Boghossian venía a decir que, en cuanto alguien como Peter no podría saber *a priori* qué argumento es válido y cuál no, entonces nuestras adscripciones *de dicto* no podrían racionalizar sus creencias y explicar su conducta. Y era *éste* el problema, que Peter es intuitivamente racional y que nuestras adscripciones no pueden explicar por qué. El problema no era que si el externismo es verdadero, entonces Peter no puede asegurar su racionalidad, sino que si el externismo es verdadero, entonces nuestras adscripciones *de dicto* no pueden explicar que Peter es racional.

Recordemos la siguiente cita que ya dábamos en el primer capítulo, en la que Boghossian explica por qué sería problemático que alguien no pueda saber *a priori* qué argumentos son válidos y cuáles no:

> ¿Para qué necesitamos la tesis de la *aprioricidad* de la lógica? (...) una consideración intuitiva: la tesis juega un papel importante en fijar qué es para

[228] But even with the addition of this identity premise, the argument still relies on the identity of concepts. In particular, it relies on the fact that the concept expressed by the first use of ['Pavarotti' in 0: "Pavarotti is Pavarotti"] is identical with the concept expressed by the use of ['Pavarotti' in 1: "I met Pavarotti at Lake Taupo"]; and, that the concept expressed by the second use of ['Pavarotti' in 0] is identical with the concept expressed by the use of ['Pavarotti' in 2: "I listened to Pavarotti yesterday"]. If one can never rely on the identity of concepts through an inference, the argument needs supplementing with further identity premises. But we have now started a regress. Any additional identity premise of the form ['Pavarotti is Pavarotti'] relies on an identity between the concept expressed by the first use of ['Pavarotti'] in this premise and that expressed by the use of ['Pavarotti'] in some other premise. If one is never allowed to rely on the identity of concepts, the inference must involve a further identity premise. But the same issue arises with any such further premise. (Brown (2004), p. 181)

> nuestras actitudes proposicionales racionalizar nuestras conclusiones prácticas y teóricas.

La estrategia de la premisa oculta afirma que explicamos correctamente la racionalidad de Peter y su conducta si le adscribimos la creencia de que Pavarotti es bi-Pavarotti (creencia que Peter sí tiene). No hay peligro de regreso en nuestra explicación porque la estrategia de la premisa oculta en ningún momento dice que no podemos basarnos en una relación de identidad: cuando nuestras creencias de identidad son verdaderas, éstas no constituyen premisas en nuestras inferencias—no necesitamos adscribirle a Peter* la creencia de que Pavarotti es Pavarotti para explicar que es racional. Si suplimos nuestra explicación con una premisa de identidad, eso no es así porque no podemos basarnos en una relación de identidad que de hecho se da, sino porque sin esa premisa la explicación no es buena.

Collins (2008) menciona un último motivo que podría tener alguien para rechazar la estrategia de la premisa oculta. Aduce que si se lo preguntáramos, Peter diría que la premisa de identidad es superflua en el argumento:

> Si el argumento, sin la premisa, fuera válido, como Peter supone que es, entonces la premisa oculta sería completamente superflua. Presionemos a Peter sobre el estatus de la premisa oculta, y, si realmente sabe de lógica, insistirá que la premisa es superflua para la validez de la inferencia, y que no es un constituyente de su inferencia. [229]

Ya hemos respondido a esta crítica en la sección anterior: preguntarle a Peter sobre si la inferencia que sigue contiene una premisa de identidad o no, simplemente no viene al caso—puede estar equivocado al respecto. Peter cree que Pavarotti es bi-Pavarotti, y esta creencia es crucial para que adopte la creencia de que una vez se encontró en el Lago Taupo al tenor que escuchó anoche. Esto es suficiente para que concluyamos que esa creencia constituye una premisa en la inferencia que sigue Peter; que él no lo crea así no demuestra nada.

Terminemos este capítulo con un ejemplo que tiene como protagonista a alguien que tenía creencias erróneas sobre qué premisas contenía una inferencia que consideraba. En

[229] If the argument, without the premise, were valid, as Peter supposes it is, then the suppressed premise would be completely superfluous. Press Peter on the status of the suppressed premise, and, if he really has training in logic, he would insist that the premise was superfluous for the validity of the inference, and that it was not a constituent of his inference. (Collins (2008), p. 569)

la época en la que escribió el *Tractatus*, Wittgenstein creía que había al menos dos personas que se adherían al programa logicista en Filosofía de la Matemática. Y basaba esta opinión (podemos suponer) en su creencia de que Bertrand Russell se adhería al programa logicista, y que así lo hacía también Gottlob Frege.

¿Creía Wittgenstein que la inferencia en la que se basaba su creencia de que había al menos dos logicistas contenía una premisa de no-identidad "Russell no es Frege"? Parece razonable pensar que no, y es que en el *Tractatus* Wittgenstein defendía que la siguiente forma argumental era válida (sin necesidad de una premisa $a \neq b$):

(1) Fa
(2) Fb
(3) $\exists x \exists y [(Fx \wedge Fy) \wedge (x \neq y)]$

A diferencia de Peter, Wittgenstein tenía incluso creencias erróneas sobre qué formas argumentales son válidas, y cuáles no. ¿Tenemos que concluir que el argumento que de hecho aceptaba Wittgenstein carecía de una premisa de no-identidad, y que en parte por eso, Wittgenstein era irracional?

No. Wittgenstein no era irracional, simplemente tenía creencias erróneas sobre algunas cuestiones de lógica formal. Y, como era racional, la inferencia que seguía sí contenía la premisa "Russell no es Frege", no importa cuán vehementemente pudiera negar él este hecho. Preguntar a Wittgenstein si su inferencia contenía la premisa "Frege no es Russell" para así evaluar su racionalidad simplemente no viene al caso (basta con constatar que de hecho Wittgenstein creía que Frege no es Russell).

Concluyendo, Peter es racional porque la inferencia que considera contiene una premisa oculta de identidad; las adscripciones *de dicto* del externista racionalizan perfectamente su situación. No hay *puzzle* que resolver, ni problema para el externista—lo único que había era una mala descripción de la situación de Peter. Además, esta respuesta surge de forma natural cuando comparamos la situación de Peter con otros ejemplos más comunes.

4. MORDER LA BALA

Las respuestas que hemos presentado hasta ahora niegan que las adscripciones *de dicto* del externista no puedan racionalizar las creencias y la conducta de Peter. Rechazan la interpretación que hace Boghossian del ejemplo: Tye niega que la primera premisa del argumento que considera Peter sea sobre Pavarotti, Schiffer y Burge que la segunda premisa sea sobre bi-Pavarotti, y nosotros, aún aceptando la interpretación que hace Boghossian del contenido de estas dos premisas, protestamos que la suya no es una buena descripción de en qué situación se encuentra Peter (porque obvia una creencia de identidad que tiene éste).

Las respuestas que presentaremos en este último capítulo del trabajo aceptan la interpretación de Boghossian, que Peter asume como válida una inferencia que de hecho no lo es. Por eso, en mayor o menor medida, "muerden la bala" lanzada por Boghossian.

Hemos concluido el primer capítulo con la siguiente cita:

> Si abandonamos la transparencia incluso para los pensamientos *de dicto*, entonces debemos o bien abandonar la noción de racionalidad y con ella la práctica de explicación psicológica que sustenta, o bien debemos demostrar que estas nociones pueden ser remodeladas de modo que no lleven a resultados absurdos. El problema es que la primera sugerencia es descabellada y no parece que haya un modo obviamente satisfactorio de implementar la segunda.

Así, diferenciaremos entre dos "modos de morder la bala". El primero acepta que las adscripciones del externista no pueden racionalizar creencias y conductas, pero responde que otros factores son relevantes para evaluar la racionalidad del sujeto. El otro simplemente acepta que, muchas veces, la reflexión *a priori* no basta para saber qué exige la racionalidad de nosotros, y admite que existe "la suerte lógica". Argumentaremos que ninguna de estas opciones es buena. Primero, tenemos dudas sobre si entienden correctamente el argumento de Boghossian. Y, segundo, nos parece que (al menos Faria y Sorensen) terminan asumiendo una posición un tanto extrema que el externista no está comprometido a adoptar.

4.1. BROWN Y UNA NOCIÓN ALTERNATIVA DE RACIONALIDAD

Brown acepta las primeras dos premisas del argumento de Boghossian; afirma que si el externismo semántico es verdadero, entonces uno no puede conocer *a priori* las propiedades lógicas de sus pensamientos e inferencias. Así, ejemplos como el de Peter son posibles; si el externismo es verdadero, se puede dar un escenario donde uno juzga como válida una inferencia que de hecho no lo es (sin que la reflexión *a priori* pueda ayudar a corregir el error).

> No importa cuán sofisticada sea la comprensión que tiene un sujeto de la lógica, si para ella la mismidad y la diferencia del contenido no son transparentes, no será capaz de conformar a priori sus pensamientos a los principios de la lógica.[230]

Nos recuerda una distinción que ya hacían Boghossian (1992a) y Campbell (1987): una cosa es poder saber *a priori* qué forma lógica ha de tener un argumento para ser válido, y otra poder saber *a priori* si los argumentos concretos que consideramos satisfacen alguna de esas formas. Brown llama "B-racionalidad" a la capacidad de adecuar *a priori* los pensamientos concretos que tenemos a las leyes de la lógica—con Boghossian, Brown acepta que el externismo mina nuestra B-racionalidad.

[230] No matter how sophisticated a subject's grasp of logic, if sameness and difference of content is not transparent to her, she will be unable to make her thoughts conform to the principles of logic a priori. (Brown (2004), p. 184)

La cuestión es que, de acuerdo con Brown, esto no supone un gran problema para el externista. Primero, nos recuerda que de hecho tendemos a no ser B-racionales. Es común que la gente cometa errores de carácter lógico, que adopte creencias que son contradictorias entre sí o argumentos que no son válidos. Dado que de hecho no somos B-racionales, "difícilmente podrá ser una objeción al anti-individualismo en particular que tiene la consecuencia de que los sujetos ordinarios no son B-racionales"[231].

Segundo, no se sigue que, como predecía Boghossian, el externista tendrá problemas para explicar conductas. Recordemos que éste mencionaba ciertas leyes que relacionaban actitudes psicológicas y conducta[232]. Decía que, si el externismo es verdadero, estos principios ya no explican la conducta de los sujetos. Ahora, Brown responde que uno puede aceptar estos principios como generalizaciones que son útiles, sin que los contraejemplos demuestren que no son generalizaciones fiables:

> Incluso si hay contraejemplos a la transparencia, podemos todavía usar los principios psicológicos descritos arriba como una guía para la predicción y explicación. En cuanto los casos en los que un sujeto tiene inadvertidamente términos sinónimos, o inadvertidamente expresa dos conceptos diferentes mediante un único término, son poco frecuentes, las generalizaciones psicológicas seguirán siendo generalmente verdaderas. Así, podemos continuar usando estos principios como base para predicciones y explicaciones, atemperados a la evidencia sobre si el sujeto es consciente de las relaciones de mismidad y diferencia que hay entre los contenidos de sus pensamientos.[233]

Por último, señala que el externista podría adentrarse en un proyecto de remodelación de la racionalidad. Así, uno podría intentar distinguir entre aquéllos que son B-irracionales e irracionales (*simpliciter*) y aquéllos que son B-irracionales pero racionales. Uno podría intentar marcar esta distinción acudiendo a las presuposiciones del argumento no-válido que acepta un sujeto. Entre esas presuposiciones estarían las creencias de (no-)identidad y co-referencia que tiene el sujeto:

---

[231] It can hardly be an objection to anti-individualism in particular that it has the consequence that ordinary subjects are not B-rational. (Brown (2004), p. 189)

[232] Si **S** cree **p** en t y tiene en t la intención de **F** si **p**, y si S no tiene razones independientes para no acometer **F**, entonces S intentará acometer **F** o, al menos, tendrá la disposición de intentar acometer **F**. Si S tiene la intención de **F** syss **p**, pero no cree **p**, sino **q** en cambio, (donde **p** y **q** son proposiciones lógicamente independientes), entonces S no intentará acometer **F**.

[233] Even if there are counterexamples to transparency, we may still be able to use the psychological principles outlined above as a guide to prediction and explanation. As long as cases in which a subject unwittingly has synonymous terms, or unwittingly expresses two different concepts by a single term, are infrequent, then the psychological generalizations will still be generally true. Thus, we can continue to use these principles as a basis for prediction and explanation tempered by evidence about whether the subject realizes the relations of sameness and difference between her thought contents. (Brown (2004), p. 190)

> El anti-individualista podría también sugerir algún modo de marcar la diferencia entre esos errores lógicos por los cuales censuramos a los agentes, y aquéllos por los cuales no lo hacemos. (...) Un modo de marcar esta diferencia sería mantener que un sujeto no es censurable por tener un par de creencias contradictorias de la forma *a* es *F* y *a* no es *F* si, al preguntársele, dudara o negara que las creencias conciernen al mismo objeto. Del mismo modo, podríamos mantener que un sujeto no es censurable por hacer una inferencia que no es válida debido a una falacia de equivocación si, al preguntársele, afirmara erróneamente la identidad relevante.[234,235]

Peter, por ejemplo, no es B-racional, porque acepta como válida una inferencia que de hecho no lo es, pero no se sigue que sea irracional. Es una presuposición de su argumento que Pavarotti es bi-Pavarotti, y esta presuposición nos permite poder diferenciarle de aquellos casos claros en los que queremos decir que el sujeto es irracional.

Por lo tanto, según Brown, sí es verdad que el externismo semántico tiene la consecuencia de que no somos B-racionales. Pero esto no supone un gran problema ya que, primero, somos B-irracionales independientemente de que el externismo sea verdadero o no y, segundo, porque podemos remodelar nuestra noción de racionalidad de tal modo que las presuposiciones de las inferencias que aceptamos delimitarán qué B-irracionales son racionales y qué B-irracionales no lo son[236].

---

[234] The anti-individualist may also suggest a way of drawing the distinction between those logical failings for which we blame agents, and those for which we do not. (...) One way to draw this distinction would be to say that a subject is blameless for having a pair of contradictory beliefs of the form *a* is *F* and *a* is not *F* if, were the question to arise, she would doubt or deny that the beliefs concern the same object. Similarly, we may say that a subject is blameless for making an inference that is invalid through a fallacy of equivocation if she would mistakenly affirm the relevant identity, were the question to be raised. (Brown (2004), pp. 190-191)

[235] Mencionemos que también Heal (1998) sugiere esta misma idea (aunque más adelante la rechaza porque no cree que la cohabitación de conceptos sea una opción aceptable): "Podemos simpatizar con las inferencias a las que llega una persona guiada por este tipo de error (como confundir PAVAROTTI con BI-PAVAROTTI), así como encontrarlas inteligibles, porque, en un pensador auto-consciente y reflexivo, las identidades entre conceptos que se asumen de seguro pueden jugar algún rol en motivar inferencias del mismo modo que las identidades actuales. (...) Sugiero que todavía no tenemos entre nuestras manos argumentos que nos permitan decir que el externismo, considerado en términos generales, debería excluir la posibilidad de que una persona cometa errores sobre la identidad de sus conceptos y por ello haga errores relacionados como aquéllos sobre la forma lógica de los argumentos que ofrece." ("We can sympathise with and find intelligible the inferences to which a person is led by this kind of mistake [i.e. mistaking PAVAROTTI with BI-PAVAROTTI], because, in a self-conscious and reflective thinker, assumed identities of concepts may surely play some role in motivating inferences as well as actual identities. (...) I suggest that we do not yet have any detailed arguments to hand in the light of which we can say that externalism, considered in general terms, ought to exclude the possibility of a person making mistakes about the identity of his or her concepts and hence making related mistakes such as those about the logical form of arguments he or she offers." (Heal (1998), pp. 104-105))

[236] Mencionemos al menos que algunos autores adscritos a la corriente neo-russelliana a la que hemos aludido en la introducción al trabajo también se han visto embarcados en un proyecto parecido de

4.2. SORENSEN Y FARIA: SUERTE LÓGICA Y RACIONALIDAD A POSTERIORI

Alguien podría simplemente negar que siempre tenemos a mano la posibilidad de ser racionales: la reflexión *a priori* no basta siempre para saber qué exigen de nosotros la lógica y la racionalidad, la investigación *a posteriori* puede marcar alguna diferencia en nuestra racionalidad. Unos pocos autores han optado por esta vía, algunos por motivos independientes a la discusión que estamos presentando en esta parte[237], otros, Sorensen (1998) y Faria (2009) por ejemplo, como respuesta al argumento de Boghossian. Estos últimos concluyen que el que uno adopte una posición lógicamente aceptable (y racional) también es en parte una cuestión de suerte, y que esto tiene consecuencias para nuestra noción de *responsabilidad lógica*. Seguramente lo más curioso de la propuesta de Sorensen y Faria sean los ecos explícitos a unos debates análogos que se dieron en Ética durante los años setenta del siglo XX[238].

Los dos, Faria y Sorensen, ofrecen algunos ejemplos que más o menos se parecen al de Peter. Faria nos sugiere que imaginemos que, un día al volver a casa, nos topamos con el pastor alemán del vecino. El perro está jugando y, nada más vernos, se nos acerca con una actitud amistosa. "Es éste un perro muy simpático", pensamos. Pasan algunos días y, de nuevo, al volver a casa, nos topamos con el pastor alemán del vecino. No para:

---

remodelación de la racionalidad. Al defender que la única aportación semántica de un nombre propio a la proposición expresada por el enunciado es el objeto al que refiere, el neo-ruselliano se encuentra con un escenario en el que, por ejemplo, uno puede asentir al enunciado 'Héspero es una estrella' pero disentir a 'Fósforo es una estrella' cuando, de acuerdo con la semántica neo-russelliana, estos enunciados expresan la misma proposición. Así, parece que es común que haya sujetos con creencias cuyos contenidos caen en contradicción, pero que nos gustaría decir que son racionales. La respuesta más popular entre aquellos que se adhieren a esta corriente es afirmar que las creencias (o las relaciones psicológicas detrás del predicado "cree que") son relaciones triádicas entre sujetos, proposiciones y modos de presentación. Así, por ejemplo, uno no será irracional aún en el case de que asienta y disienta a una misma proposición, si en esas dos actitudes la proposición se le presenta bajo dos modos distintos. Esto es, la adscripción *de dicto* de contenidos no basta para evaluar la racionalidad de un sujeto, es necesario tener en cuenta bajo qué modos se le presentan esos contenidos al sujeto. Véanse, por ejemplo: Crimmins y Perry (1989), Perry (1977, 1979, 1988), Soames (1988, 2002) y, sobre todo, Salmon (1986, 1989).

[237] Probablemente Williamson (2000) sea el ejemplo paradigmático.

[238] Véanse, sobre todo, Williams (1976) y Nagel (1976)

persigue palomas, hojas que caen del árbol, coches... sin dejar de correr. "Vaya, éste es un perro infatigable", pensamos.

> ¿Estoy en posición de inferir que hay un perro en mi vecindario que es ambas cosas, simpático e infatigable? Bueno, supongamos que mi vecino es un amaestrador de pastores alemanes, y que lo que he identificado exitosamente en esas dos ocasiones era un par de hermanos de la misma camada – llamémosles Harry el Simpático y Barry el Infatigable. Tal y como son las cosas, Harry no es nada infatigable, mientras que Barry es más bien antipático. Supongamos además que no hay más perros en el vecindario. Así mi conclusión es simplemente falsa, y mi razonamiento no es bueno – una clara falacia de equivocación. Su forma no es 1-3 [1. Fa; 2. Ga; 3. ∃x (Fx ∧ Gx)], sino: 4-6 [4. Fa; 5. Gb; 6. ∃x (Fx ∧ Gx)].[239]

Sorensen (1998) nos presenta a Mr. Move, quien juzga "Hace calor aquí; Está húmedo aquí; Por lo tanto, hace calor y está húmedo aquí" sin percatarse de que ha cambiado de lugar entre su pensar de la primera premisa y el de la segunda. Así, su inferencia tampoco es válida, ya que las distintas proferencias de 'aquí' refieren a distintos lugares. El uso de términos indéxicos conlleva la posibilidad del error *extrínseco*, errores de los que el sujeto no puede percatarse mediante introspección; por contra, un error lógico será intrínseco si la reflexión *a priori* basta para percatarse de él.

Así, uno puede evitar cometer errores lógicos intrínsecos, pero no está siempre en nuestras manos esquivar los errores extrínsecos. Todos los ejemplos propuestos (Peter, Leire, el pastor alemán o Mr. Move) son ejemplos de errores de este último tipo. La posibilidad de este tipo de errores demuestra que existe la "suerte lógica"; un razonamiento *a priori* impecable no basta para conseguir que nuestras inferencias se adecuen a las normas de la lógica, esto también depende de factores que nos son externos, es en parte una cuestión de suerte.

Cabría preguntarse si, del mismo modo, ser racional es también una cuestión de suerte. Por supuesto, si con Boghossian aceptáramos que ser B-racional es un requisito necesario para poder ser considerado racional, se seguiría que lo es. Por otro lado, como hemos visto en las secciones anteriores, uno podría intentar reformular nuestra noción

[239] Am I now entitled to infer that there is a dog in my neighborhood who is both friendly and restless? Well, suppose my neighbor is a breeder of Golden Retrievers and what I successively spotted on those two occasions were a pair of siblings from the same litter – call them Harry the Friendly and Barry the Restless. As things go, Harry is not excitable at all, while Barry is of a rather unfriendly disposition. Suppose further there are no other dogs in the neighborhood. So my conclusion is just false, and my reasoning unsound – a plain fallacy of equivocation. Its form is not 1-3 [1. Fa; 2. Ga; 3. ∃x (Fx ∧ Gx)] but rather: 4-6 [4. Fa; 5. Gb; 6. ∃x (Fx ∧ Gx)]. (Faria (2009), pp. 4-5)

de racionalidad, o intentar explicar la racionalidad de un sujeto acudiendo a elementos como el contenido *estrecho*. Sorensen tiene clara su apuesta:

> El externista tendría que sentirse inclinado a adoptar una posición más social, una que no confine las evaluaciones de racionalidad a la psicología estrecha del sujeto.[240]
>
> No obstante, la validez de las inferencias es una parte central de la racionalidad. Por eso, hay irracionalidades extrínsecas.[241]

Ser racionales es en parte una cuestión de suerte, no depende completamente de nosotros. La reflexión *a priori* no basta para cumplir con las exigencias de la racionalidad, algunas veces necesitamos de investigación *a posteriori* para conseguir ser racionales. La cuestión es que, según Sorensen, dado que toda teoría semántica está comprometida a la existencia de términos indéxicos, y dado que la existencia de términos indéxicos abre la puerta a errores lógicos extrínsecos, difícilmente podría ser un problema para el externista el estar comprometido a la suerte lógica y a la racionalidad *a posteriori*.

Abramos aquí un pequeño paréntesis y expliquemos por qué decimos nosotros que Brown (2004) se adentra en un proyecto de remodelación de la racionalidad pero no así Sorensen (1998) y Faria (2009)—y es que uno podría entender que las afirmaciones de Sorensen y Faria suponen también una remodelación de la noción tradicional de racionalidad. Según esta noción tradicional, primero, uno es racional sólo si es B-racional y, segundo, la reflexión *a priori* le basta a uno para ser B-racional y racional. Tanto Brown (2004) como Sorensen (1998) y Faria (2009) niegan que uno pueda mantener estas dos ideas, ya que niegan que la reflexión *a priori* basta para ser B-racional. En este sentido, es lícito decir que los tres rechazan el modelo tradicional de racionalidad (y, por lo tanto, se embarcan en un proyecto de remodelación de esta noción).

Pero hay una diferencia entre Brown, por un lado, y Faria y Sorensen, por el otro. Brown asume que no somos B-racionales, pero propone que es posible distinguir entre los B-irracionales que son racionales y los que no lo son. Además, sugiere que esta

[240] An externalist should be inclined to take a more social stand, one that does not confine rationality assessments to the individual's narrow psychology. (Sorensen (1998), p. 330)
[241] Nevertheless the validity of inferences is a central part of rationality. Thus there are extrinsic irrationalities (Sorensen (1998), p. 331)

distinción permite que la reflexión *a priori* baste a uno para ser racional: es cierto que la reflexión *a priori* basta para ser racional, aunque "racional" no se entiende como se ha hecho tradicionalmente—por eso decimos que Brown se embarca en un proyecto de remodelación de la racionalidad.

Por contra, parece que Sorensen y Faria tienden a seguir la vía alternativa. Asumen que uno es racional sólo si es B-racional; así, como hay posibilidad de error lógico extrínseco, es falso que la reflexión *a priori* siempre sea suficiente para asegurar la racionalidad de uno. Dado que, como hemos dicho, la idea de que la reflexión *a priori* basta para ser racional le es central a la noción tradicional de racionalidad, uno podría sugerir que también Sorensen y Faria remodelan nuestra noción de racionalidad. Pero, a diferencia de Brown, éstos no discrepan con el modelo tradicional sobre qué condiciones ha de cumplir uno para poder ser considerado racional, lo que niegan es que uno pueda cumplir con ellas *a priori*. Es en este sentido que decimos que Sorensen y Faria no proponen una noción alternativa de racionalidad.

Volvamos con Sorensen y Faria. Como hemos dicho, sugieren una analogía entre estas discusiones y las que se dieron sobre la existencia o no de la "suerte moral". Ahora bien, normalmente reprochamos al irracional que lo es; ¿se sigue que deberíamos condenar a los protagonistas de los ejemplos que hemos tratado por ser irracionales?

Faria y Sorensen adoptan posiciones distintas en esta cuestión. Faria acepta la máxima de que "*deber* implica *poder*"[242] y, por eso, afirma que sólo es condenable aquél que podría haber evitado su error:

> En este tipo de ejemplos más comunes [el ejemplo de Leire o el del pastor alemán], la información de la cual de hecho carece el sujeto *está* a su disposición: el sujeto la obtendría con sólo preocuparse lo suficiente como para conocerlo. No sucede así en las transiciones lentas – de allí las posiciones exculpatorias de las que me quejaba [la estrategia basada en reemplazo conceptual, la estrategia Schiffer-Burge, y la estrategia de la premisa oculta].[243]

Esto es, hay diferencias entre las ficciones al estilo de las transiciones lentas y los ejemplos "más mundanos". Sabiendo que 'Pepe' es un nombre bastante común, Leire

[242] "*Ought* implies *can*", en inglés.
[243] In such down to earth cases, the information which the subject actually lacks *is* available: the subject would be apprised of it if only she cared enough to know. Not so on the slow switching scenarios – hence the exculpating moves of which I was complaining. (Faria (2009), p. 13)

podría preguntar si aquel Pepe que escribe artículos de semántica es el mismo Pepe que canta en un grupo de pop-rock, aquél que razona sobre los pastores alemanes del vecino tiene a mano investigar si se trata del mismo animal o no, pero a Peter le es absolutamente imposible descubrir que Pavarotti no es bi-Pavarotti. Tal y como se construyen las transiciones lentas, al sujeto le es imposible adquirir la información que le permitiría evitar su error. Esto ha motivado, de acuerdo con Faria, que varios autores hayan intentado salvar la racionalidad de Peter (y así, la de los protagonistas de los ejemplos más mundanos), lo cual es un error según él—Peter es irracional, aunque no se sigue que esto le sea reprochable. Ahora bien, en los ejemplos más normales, uno puede adquirir la información que le evitaría su error, y por eso, sí tiene cierta responsabilidad en aceptar la inferencia que acepta. La norma "*deber* implica *poder*" nos muestra quién merece censura y quién no.

Sorensen parece rechazar esta norma de que "*deber* implica *poder*"

> Las equivocaciones extrínsecas (...) reducen la fiabilidad de los razonamientos de uno. Deberíamos esperar que los demás desaprobaran estas equivocaciones lingüísticas, porque las conclusiones poco fiables tienden a pasarse de unos a otros. La crítica es castigo.[244]

> Así, las creencias contradictorias deberían dividirse en tres clases. Primero está la clase familiar de las contradicciones que se pueden evitar. Luego están las contradicciones que se consideran irreprochables por su inevitabilidad. La última clase consiste de contradicciones inevitables que, a pesar de ello, condenamos – instancias en lógica de responsabilidad estricta.[245]

El mal razonamiento de un individuo disminuye la fiabilidad de los razonamientos que se llevan a cabo en la comunidad (aunque se den debido a errores extrínsecos), en principio esto sugiere que la comunidad debería recriminar al individuo su irracionalidad (aun cuando es extrínseca). La cuestión es que, según Sorensen, cada sociedad marca sus propias políticas de castigo; uno puede en una comunidad ser considerado responsable de una inferencia inválida que ha aceptado, pero no así en otra. Así, podemos distinguir entre errores lógicos inevitables que condenamos y errores

[244] Extrinsic equivocations (...) lower the reliability of one's reasoning. We should expect others to disapprove of these linguistic misalignments because unreliable conclusions tend to be passed along to others. Criticism is punishment. (Sorensen (1998), p. 330)

[245] Hence contradictory beliefs should fall into three classes. First is the familiar class of avoidable contradictions. Second are contradictions that are rendered guiltless by their unavoidability. The final class consists of unavoidable contradictions that we nevertheless condemn – instances of strict liability in logic. (Sorensen (1998), p. 332)

lógicos inevitables que no condenamos, y asumir que cada comunidad marca qué errores caen en un grupo y cuáles en el otro.

Resumiendo, Faria (2009) y Sorensen (1998) asumen que existe el error lógico extrínseco. La reflexión *a priori* no basta para adecuar las creencias de uno a las normas de la lógica y, por eso, tampoco para asegurar la racionalidad de uno. Ser racional es una cuestión de suerte. Pero niegan que esto sea un problema; es una consecuencia natural de nuestras teorías semánticas.

4.3. INTERPRETACIÓN DEL ARGUMENTO Y EVALUACIÓN DE LAS PROPUESTAS

Terminemos. Opinamos que las propuestas que hemos descrito en este último capítulo no suponen una respuesta ni satisfactoria ni atractiva al argumento expuesto por Boghossian.

Primero, tenemos dudas acerca de en qué medida estas posiciones suponen una respuesta al argumento. Hemos caracterizado las opciones presentadas en este capítulo como alternativas que "muerden la bala" lanzada por Boghossian; pero es importante fijarse en cuál es la bala que muerden. Brown (2004) asumía que es verdad que el externismo tiene la consecuencia de que no somos B-racionales, pero respondía que de hecho no lo somos, y que por eso esto no supone un grave problema para el externista. La respuesta de Sorensen (1998) y Faria (2009) viene a decir que muchas veces necesitamos de la investigación *a posteriori* para saber qué argumento es válido y cuál no: existen los errores lógicos extrínsecos y la suerte lógica.

Creemos que estas respuestas sugieren que Brown, Sorensen y Faria entienden que el problema que pretende traer a colación Boghossian es el siguiente. Si el externismo es verdadero, entonces Peter acepta como válido un argumento que de hecho no lo es y, por lo tanto, no es racional. Pero esto es un problema, porque la reflexión *a priori* no le bastará para corregir su error, y se supone que es parte de la noción tradicional de

racionalidad que la reflexión *a priori* le es suficiente a uno para ser racional. Esto es, el problema es que si el externismo es verdadero, entonces Peter no puede "asegurarse" *a priori* su racionalidad.

A esto Brown le responde dos cosas; primero, que no es un problema para el externista que tenga que decir que Peter no es B-racional, porque de hecho no lo somos y, segundo, que deberíamos remodelar nuestra noción tradicional de racionalidad para que Peter y otros B-irracionales puedan ser considerados racionales. Sorensen y Faria le responden que Peter es racional, pero que esto no es un problema para el externista, porque de hecho existen los errores lógicos externos y, por eso, deberíamos rechazar la noción tradicional de racionalidad según la cual la reflexión *a priori* basta para ser racional.

La cuestión es que nos parece que no era ése exactamente el problema que pretendía identificar Boghossian. Brevemente, el problema no era que si el externismo es verdadero, entonces Peter no puede "asegurar su racionalidad", sino que queremos decir que Peter es racional, y el externista no tiene las herramientas necesarias para decir que lo es. Peter actúa de cierta manera, entre otras cosas, profiere ciertos enunciados. Queremos explicar por qué lo hace y hacemos esto, en parte, adscribiéndole creencias y otras actitudes que explican que es racional. Lo que viene a decir Boghossian es, creemos, que el externista no puede explicar que Peter es racional, que carece de las herramientas necesarias para hacerlo (sus adscripciones *de dicto* no explican por qué actúa Peter del modo en que lo hace, por qué profiere los enunciados que profiere, o por que tiene las disposiciones que tiene). Y no vemos que responder que de hecho somos B-irracionales o que de hecho cometemos errores lógicos externos sea una respuesta a este problema.

Decíamos que Brown respondía a Boghossian que de hecho no somos B-racionales, que hay varios ejemplos que lo atestiguan. Por ejemplo, es común que cometamos errores sobre la validez de argumentos muy simples. Pero hay una diferencia importante entre Peter y los ejemplos de B-irracionalidad que menciona Brown: los sujetos de sus ejemplos podrían corregir su error mediante reflexión *a priori*, Peter no (a diferencia de los sujetos B-irracionales de Brown, Peter *está condenado a ser B-irracional*). Boghossian tiene una respuesta fácil contra Brown: a diferencia de Peter, esos sujetos

que menciona no son *intuitivamente* racionales, porque cometen errores de carácter lógico que podrían haber evitado. Acudir a estos ejemplos de poco servirá, pues, para responder a Boghossian; decir que de hecho no somos B-racionales no ayuda a explicar que Peter es racional.

Sí que es verdad que de las propuestas de Sorensen y Faria se sigue una respuesta al argumento de Boghossian. Éste decía que el externista tenía un problema, porque no podía explicar que Peter fuera racional, y queremos decir que lo es. Parece que Sorensen y Faria simplemente responderían que Peter no es racional.

No creemos que ésta sea una buena respuesta, Sorensen y Faria están proponiendo una posición incómoda que el externista no tiene por qué adoptar. Realmente queremos decir que Peter tiene motivos para adoptar la creencia de que una vez se encontró en el Lago Taupo con el tenor que escuchó cantar anoche, que está justificado al hacerlo; es esto lo que queremos decir al afirmar que Peter es intuitivamente racional, algo se perdería si nuestras adscripciones *de dicto* no pudieran explicar que lo es.

Por supuesto uno puede responder que Peter no tiene justificación para creer que el tenor que vio en el Lago Taupo es el tenor que escuchó anoche, que no debería adoptar esta creencia, pero esta posición deja de lado por completo algunas de nuestras intuiciones sobre en qué situación tiene uno justificación para creer que *p*. El externista estaría cediendo algo importante si adoptara esta posición, tal que esto supondría una clara desventaja en la discusión que mantiene con el internista. Además, es ésta una alternativa a la que el externista no tiene por qué acudir, ya hemos explicado en el tercer capítulo que tiene a mano una buena explicación de cómo es que Peter es racional y hace bien en adoptar la creencia en cuestión.

Mencionemos para acabar que Brown (2004) sí abre la puerta a una posición que puede ser atractiva. Brown respondía que, aunque uno acepte como válida una inferencia que no lo es, puede ser considerado racional dependiendo de cuáles son las presuposiciones de esa inferencia; entre esas presuposiciones mencionaba las creencias de (no-)identidad del sujeto. Sobre esta base sí se puede desarrollar una respuesta satisfactoria al argumento de Boghossian; las adscripciones *de dicto* del externista sí racionalizan la

posición de Peter (a pesar de que no todas las creencias adscritas constituyen premisas en su inferencia, y de que esto le lleva a mantener que Peter no es B-racional).

La cuestión es que no vemos diferencias importantes entre esta posición y la estrategia de la premisa oculta que hemos defendido en el capítulo anterior. La única diferencia entre ellas reside en que la estrategia de la presuposición no dona categoría de premisa a las creencias de (no-)identidad del sujeto. Y no vemos qué puede determinar que una creencia que tiene uno y que motiva que siga una inferencia sea una presuposición y no una premisa de esa inferencia. Peter por ejemplo llega a la creencia de que el tenor que vio en el Lago Taupo es el mismo que escuchó cantar anoche, y lo hace infiriéndolo de otras creencias que tiene. La creencia de que una vez vio a Pavarotti en el Lago Taupo es relevante para que Peter acepte la inferencia en cuestión, pero también lo es su creencia de que Pavarotti es bi-Pavarotti—no vemos qué puede decidir que la primera constituye una premisa en la inferencia pero no así la segunda.

Además, mantener que la premisa de identidad constituye una premisa en la inferencia de Peter nos permite afirmar que éste es B-racional. A diferencia de la estrategia de la presuposición, la estrategia de la premisa oculta puede seguir manteniendo la noción tradicional de racionalidad, que afirma que uno es racional sólo si es B-racional. Por eso, la estrategia de la premisa oculta parece que compromete menos al externista. Puede que esto mueva un poco la balanza en favor de la propuesta que hemos defendido nosotros pero, de nuevo, no creemos que haya grandes diferencias entre ésta y la respuesta afirmando que podemos explicar la racionalidad de Peter si acudimos a las presuposiciones de identidad de sus inferencias.

5. ÚLTIMOS COMENTARIOS Y CONCLUSIONES

Concluyamos esta tercera y última parte del trabajo. A lo largo de esta parte hemos presentado y discutido un tercer argumento anti-externista propuesto por Boghossian (1992a, 1994). Según el argumento, si el externismo semántico fuera verdadero, entonces la víctima de un transición lenta aceptaría como válidas algunas inferencias que de hecho no lo son, y toda la reflexión *a priori* no le serviría para poder corregir su error. Esto es problemático porque, primero, querríamos decir que alguien en esta situación es racional, y parece que uno no lo es si acepta como válidas inferencias que no lo son. Por otro lado, esto mostraría cierta tensión entre el externismo semántico y los motivos por los cuales hacemos adscripciones de actitudes *de dicto*; es una finalidad de nuestras adscripciones racionalizar las creencias y explicar la conducta de los sujetos de las adscripciones, y el objetivo del argumento es demostrar que el externismo no es compatible con que nuestras adscripciones cumplan estos cometidos. Además, según Boghossian, el externismo tiene esta consecuencia porque es incompatible con la tesis de que el contenido es transparente.

A lo largo de esta parte hemos presentado y discutido algunas posibles respuestas que tiene a mano el externista. En el segundo capítulo hemos estudiado dos propuestas que discrepan con Boghossian sobre qué creencias constituyen las premisas de una inferencia que considera la víctima de una transición lenta; de acuerdo con estas

propuestas, cuando alguien en esta situación considera una inferencia deductiva, no corre el peligro de confundir entre sí los conceptos que forman la inferencia. Unos apelan a pérdidas y reemplazos conceptuales (afirmando que el externismo es compatible con la transparencia del contenido), otros a referencias anafóricas implícitas.

Pero ninguna de estas dos opciones nos convence. Por un lado, ya en la segunda parte hemos dejado claro que apostamos por un modelo de cohabitación conceptual en detrimento de uno de reemplazo, y hasta donde llegamos a ver, el externismo será compatible con la transparencia del contenido sólo si asumimos que en las transiciones se da reemplazo. Por otro lado, nos hemos adherido a la afirmación hecha por Collins (2008) de que el defensor de la anáfora implícita no llega a explicar cómo alguien como Peter llega a creencias de bi-Pavarotti partiendo de sus creencias de Pavarotti; la estrategia Schiffer-Burge es una "no-respuesta" al problema planteado por Boghossian.

En el cuarto capítulo hemos estudiado dos alternativas que proponen que el externista puede "morder la bala", asumir en cierto modo las consecuencias que le echaba en cara Boghossian. Así, quienes se apuntan a esta estrategia proponen remodelar en parte nuestra noción tradicional de racionalidad: algunos responden que aceptar inferencias que de hecho no son válidas no es motivo suficiente para ser irracional (Brown (2004)), otros que la investigación *a posteriori* puede marcar alguna diferencia en nuestra racionalidad (Sorensen (1998) y Faria (2009)).

Hemos manifestado nuestras dudas sobre si estas respuestas entienden correctamente el argumento. Éste viene a decir que las adscripciones *de dicto* del externista no pueden explicar que alguien como Peter es racional, y no viene a ser una buena alternativa responder que muchas veces no somos racionales (que es lo que en parte hace, por ejemplo, Brown (2004)). Por otro lado, hemos criticado que la posición de Sorensen y Faria es bastante extrema (no es compatible con nuestras intuiciones de cuándo está uno justificado a la hora de adoptar una creencia), y que el externista no tiene ningún motivo para asumirla; por suerte, el externismo no tiene estas consecuencias.

Brevemente, en esta tercera parte hemos llegado a dos conclusiones principales. La primera es que el argumento de Boghossian no plantea ningún problema importante

para nadie, la segunda que la tesis de la transparencia del contenido no juega los roles tan importantes que le suponen algunos.

En el tercer capítulo hemos presentado nuestra propuesta, que hemos llamado *la estrategia de la premisa oculta*, que viene a decir que en las inferencias supuestamente problemáticas hay premisas de identidad ocultas que salvan la validez del argumento. Así, el argumento de Boghossian no supone un gran problema para el externista, ya que se basaba en una descripción del todo inapropiada de los ejemplos supuestamente problemáticos. La interpretación que hacía Boghossian de los ejemplos obviaba una creencia de identidad que tenían los protagonistas de esos ejemplos, y que era necesaria para poder explicar su racionalidad y su conducta.

Se sigue que estos escenarios no suponen ningún problema para el externista—éste sí tiene a mano adscripciones *de dicto* que racionalizan las creencias y la conducta del individuo en cuestión. No había ni *puzzle* ni problema, sólo una mala descripción de un escenario. Además, hemos argumentado, ésta, una buena respuesta al supuesto problema expuesto por Boghossian, es la "más natural" de entre las respuestas que tenemos a mano, ya que está en consonancia con las prácticas de racionalización y explicación que seguimos con respecto a otros ejemplos parecidos.

Hemos mencionado también que esta estrategia tiene una consecuencia que a algunos les puede parecer extraña, pero que nosotros creemos que el externista tendría que aceptar. Si la estrategia de la premisa oculta tal y como la hemos descrito nosotros es verdadera, entonces algunos factores externos a un sujeto como, por ejemplo, la veracidad o falsedad de sus creencias de identidad, pueden ser relevantes a la hora de determinar si una inferencia concreta que sigue contiene una premisa de identidad o no. Las propiedades internas de un sujeto no bastan para determinar qué premisas tienen las inferencias que sigue.

La segunda conclusión concierne a la transparencia del contenido. Boghossian pretendía que su argumento demostrara el rol tan importante que juega esta tesis en nuestras prácticas de adscripciones de creencias y de evaluaciones de racionalidad; nuestra respuesta al argumento demuestra que la transparencia no tiene este papel tan importante.

Hemos defendido que el externista puede explicar que alguien como Peter es racional si acude a sus creencias de identidad. Esto es, no es necesario, para poder explicar que alguien es racional, que los juicios de éste sobre las relaciones de mismidad y diferencia entre sus pensamientos y conceptos sean verdaderos, basta con incorporar estas creencias en nuestra explicación de su situación.

Por eso, es falso que la transparencia del contenido juegue un rol importante en nuestras prácticas de adscripciones de creencias y de evaluaciones de racionalidad. En la primera parte vimos que uno puede defender que tenemos auto-conocimiento autoritativo sin afirmar que el contenido es transparente. Quizás la idea de que el contenido es transparente sí tiene cierta fuerza intuitiva, pero está lejos de ser una tesis importante para explicar nada—quizás no es más que el último residuo de cierta Teoría de la Mente que creíamos haber abandonado.

(Conclusiones)

Recapitulemos. Hemos comenzado con un capítulo introductorio en el que esbozábamos el modelo externista, objeto de los argumentos que hemos estudiado a lo largo del trabajo, identificando la tesis esencial a este modelo y enumerando algunas opiniones a las que comúnmente se adhieren aquellos autores adscritos a esta corriente. Hemos presentado con detalle luego cada uno de los argumentos anti-externistas que queríamos estudiar, y caracterizado con detenimiento las que creemos les son las respuestas más destacables. Concluyamos este trabajo recordando las principales conclusiones a las que hemos llegado.

La primera conclusión que tenemos que mencionar es que ninguno de los tres argumentos consigue su objetivo: el externismo semántico es compatible con el auto-conocimiento autoritativo, y no tiene consecuencias reseñables ni para nuestras prácticas de adscripciones de actitudes proposicionales ni para nuestras explicaciones racionalizadoras.

Hemos dedicado la primera parte del trabajo a lo que podríamos llamar el argumento incompatibilista *estándar*. Hemos comenzado describiendo con detalle dos versiones alternativas de este argumento, la primera debida a Paul Boghossian, la segunda a

Jessica Brown, y hemos visto que asumen, por un lado, principios epistémicos sobre evidencia y discriminación del estilo de (RA) y (Discp) y, por el otro, una noción internista de la evidencia, tal que dos individuos tendrán exactamente la misma evidencia si se encuentran en la misma situación interna. Basándonos en las opiniones de Falvey y Owens y McLaughlin y Tye, hemos propuesto que hay cierta tensión entre estas asunciones; así, hemos identificado lo que hemos llamado *una disyunción incómoda*: o bien individuamos la evidencia externamente o bien desechamos la idea de que principios como (RA) son verdaderos (entendidos como poniendo condiciones a toda instancia de conocimiento). Una vez apostamos por una de las dos opciones ofrecidas en la disyunción, rechazaremos alguna de las premisas o presuposiciones del argumento incompatibilista.

En la segunda parte hemos estudiado otro argumento incompatibilista propuesto por Boghossian: el argumento de la memoria. Este argumento se basaba en un ejemplo de transición lenta donde alguien intentaba recordar un pensamiento que tuvo en un entorno anterior, y explotaba el supuesto compromiso del externista a negar que alguien en esa situación podría recordar. Siguiendo a Ludlow, hemos caracterizado el argumento del siguiente modo:

(1) Si S no olvida nada, entonces lo que sabe S en t_1, sabe S en t_2,
(2) S no olvidó nada,
(3) S no sabe que P en t_2;
(4) Por lo tanto, S no sabía que P en t_1.

Se supone que (1) es una obviedad, y que (2) es simplemente estipulable. El problema es que el argumento no puede ser exitoso, ya que se basa en un uso ambiguo de 'olvidar'. Tal y como se usa el término en la premisa (1), saber que *p* en t_1 y no saber que *p* en t_2 es condición suficiente para haber olvidado que *p* entre t_1 y t_2; tal y como se usa en (2), parece que es una condición necesaria para olvidar que *p* que algo haya fallado en el proceso cognitivo encargado de guardar la proposición de que *p*. Pero si saber que *p* en t_1 y no saberlo en t_2 es condición suficiente para olvidar que *p*, entonces, cuando describimos un ejemplo de transición lenta, no podemos simplemente *estipular* que el protagonista del ejemplo no olvidó nada; y, si para que S olvide que *p* es necesario que algo haya fallado en sus procesos cognitivos encargados de guardar esa

información, entonces es simplemente falso que si S sabe que *p* en t_1 y no en t_2, entonces S ha olvidado que *p* (otros factores pueden tomar parte en esa pérdida de conocimiento). Una vez desambiguamos el término 'olvidar', vemos que si (1) es verdadero, (2) no es simplemente estipulable—el argumento no puede demostrar que el externismo resulta incompatible con el auto-conocimiento autoritativo.

El argumento estudiado en la tercera parte, también debido a Boghossian, se centraba en un ejemplo de transición lenta, pero no tenía como objetivo demostrar que el externismo resulta incompatible con el auto-conocimiento autoritativo. El ejemplo nos presentaba a Peter, amante de la ópera, quien había entrado en contacto con dos tenores distintos que confundía entre sí (Pavarotti y bi-Pavarotti), y describía una situación en la que Peter parecía aceptar como válida una inferencia que de hecho no lo es, sin que la reflexión *a priori* le bastara para corregir su error. El argumento concluía que esto supone un problema grave para el externista, ya que tiene la consecuencia de que sus adscripciones *de dicto* no podrán explicar que Peter es racional, ni por qué actuaba del modo en que lo hacía. Además, Boghossian mantenía que era el supuesto compromiso del externista a negar la transparencia del contenido lo que estaba en el origen del equívoco de Peter, y alegaba que el argumento hacía patente el rol tan importante que juega esta tesis en nuestras adscripciones *de dicto* y nuestras evaluaciones de racionalidad.

Pero hemos visto que el argumento no consigue sus objetivos, porque la descripción que hace Boghossian de la situación de Peter es del todo desafortunada. Al explicar cómo llegaba Peter a las consecuencias de sus inferencias, Boghossian no tenía en cuenta una creencia que tenía Peter, a saber, que Pavarotti es bi-Pavarotti, y hemos visto que, una vez introducimos esta creencia en nuestra explicación, nuestras adscripciones *de dicto* explican a la perfección que Peter es racional. Es falso que Peter acepte como válida una inferencia que de hecho no lo es y, por eso, el argumento de Boghossian no demuestra que el externista tenga ningún problema con sus adscripciones *de dicto*, ni que la tesis de la transparencia del contenido juegue ningún rol importante en nuestras explicaciones de racionalidad y de conducta (porque la solución que le hemos dado al problema es compatible con afirmar que el contenido no es transparente).

El externista, pues, puede estar tranquilo. Pero ya dijimos en el prefacio que, más que el veredicto sobre el éxito o no de los argumentos, nos interesaban las tesis y las

discusiones que surgían en el camino, la posición que vamos bosquejando. Mencionemos, ya para acabar el trabajo, algunas de las conclusiones más importantes a las que hemos llegado a lo largo del trabajo.

Seguramente, la más importante concierne a la supuesta transparencia del contenido. Mencionamos en la introducción que en principio parecía que podría haber cierta tensión entre la teoría externista y esta tesis de la transparencia del contenido. A lo largo del trabajo hemos visto que algunos autores acuden a este supuesto compromiso del externista al construir sus argumentos anti-externistas: este supuesto compromiso constituía una premisa así en la versión de Brown del primer argumento incompatibilista como en el argumento de Boghossian sobre racionalidad e inferencia. Pero también hemos visto que algunos autores de corte externista afirman que esta teoría es compatible con mantener que el contenido es transparente, y que acuden a la idea de que en las transiciones lentas hay reemplazo conceptual para defender tal cosa.

El externista debería rechazar que el contenido es transparente, pero esto no supone ningún problema serio para él—así podríamos resumir nuestras opiniones sobre estas cuestiones.

En cuanto el externismo es verdadero, la tesis de la transparencia del contenido es falsa (al menos en cuanto a la transparencia de diferencia de contenido se refiere). Esta tesis nos dice que siempre estamos en posición de discriminar entre cualesquiera dos de nuestros pensamientos o conceptos, de saber si tienen o no el mismo contenido. En cuanto hemos abogado por un modelo de cohabitación, hemos juzgado que es posible que alguien tenga en un mismo momento dos conceptos distintos C y C', sin que sea capaz de discriminar entre ellos, porque no está en condición de activar la creencia de que C no es C' (no vemos qué podría justificar esa creencia). Si apuesta por la cohabitación de conceptos, el externista debería rechazar que el contenido es transparente.

Pero más importante que la veracidad o falsedad de la tesis de la transparencia es que ésta no juega el papel tan importante que le han otorgado algunos. Por un lado, negar que el contenido es transparente es compatible con afirmar que tenemos acceso privilegiado a nuestros estados mentales—uno ha de acudir a principios como (Discp)

para poder llegar a conclusiones incompatibilistas partiendo del compromiso del externista a negar que el contenido es transparente. Por otro lado, es simplemente falso que esta tesis juegue algún rol importante en nuestras adscripciones *de dicto* de actitudes proposicionales y en nuestras evaluaciones y explicaciones racionalizadoras (como pretendía Boghossian). La respuesta al argumento anti-externista que hemos propuesta en la tercera parte asume que el contenido no es transparente; no necesitamos que uno haga juicios correctos de mismidad y diferencia de contenidos y pensamientos para poder decir que es racional y explicar su conducta, basta con tener en cuenta cuáles son las creencias de mismidad y diferencia de contenido que de hecho tiene (aún en el caso en el que estas creencias sean falsas).

En nuestra opinión, la tesis de la transparencia del contenido es, seguramente, el último resto de un modelo de la mente que creíamos haber abandonado.

Por supuesto, nuestras opiniones sobre la incompatibilidad entre una semántica externista y la tesis de la transparencia del contenido dependen de nuestra adhesión a la idea de que en las transiciones lentas hay cohabitación de conceptos—recordemos nuestras opiniones sobre esta cuestión.

Hemos visto que estos ejemplos de transiciones se pueden interpretar de dos maneras distintas. Según la primera, en las transiciones hay *cohabitación de conceptos*, esto es, uno no pierde ningún concepto antiguo cuando adquiere el nuevo concepto en cuestión (por ejemplo, Oscar o Sally no perderían su antiguo concepto AGUA al adquirir el concepto BI-AGUA). Por contra, de acuerdo con la segunda manera de interpretar las transiciones, en éstas hay *reemplazo conceptual* (Oscar y Sally perderían el concepto AGUA al adquirir el concepto BI-AGUA).

Acudir al reemplazo conceptual puede ser tentador. Primero, porque abogar por esta idea nos permite compatibilizar el externismo semántico con la transparencia del contenido—y, se supone, algo se gana con esto (aunque nosotros no veamos qué). Además, y en parte porque posibilita al externista aferrarse a la idea de que el contenido es transparente, uno podría intentar apelar al reemplazo conceptual para responder a algunos argumentos anti-externistas—en la primera y en la tercera parte hemos visto dos ejemplos de este tipo de estrategia.

Pero estas supuestas ventajas no son tal. Primero, ya hemos dicho que no necesitamos de la tesis de la transparencia del contenido, y que podemos responder a los argumentos anti-externistas sin apelar a ella. Además, la respuesta basada en reemplazo al argumento incompatibilista de la primera parte no es buena. Como hemos visto, esta respuesta explota la idea de que los escenarios alternativos que podrían minar nuestro auto-conocimiento no son relevantes. El problema es que no dice nada sobre el auto-conocimiento de pensamientos singulares basados en ostensión, y uno debería decir algo sobre estos casos. Como hemos visto, uno podría pretender morder la bala y decir que en un escenario tal uno no sabe qué está pensando (Brown (2004) defiende esta alternativa), pero esta posición tiene la consecuencia (a nuestro juicio, extraña) de que no podemos saber que esta hipótesis es verdadera.

Así, apostar por el reemplazo no supone ninguna ventaja evidente y, además, hemos defendido que hay algunos motivos para preferir la cohabitación. Primero, porque el reemplazo da resultados extraños cuando se combina con la tesis del predominio de las transiciones lentas, la afirmación de que en el mundo actual *hay* transiciones de este tipo. Por otro lado, parece que no está claro qué podría decir el defensor del reemplazo sobre la memoria episódica (esa parte de nuestra memoria que nos provee con información sobre nuestras vivencias pasadas); si, cuando ya se ha completado la transición, Sally carece del concepto AGUA, no podrá recordar ninguna de las "vivencias de agua" que tuvo antes de sufrir la transición—parece que esto reventaría la fiabilidad de su memoria episódica. Además, no está claro qué podría motivar que alguien pierda un concepto cuando entra en contacto con una nueva sustancia, individuo o comunidad lingüística. Como dice Heal, parece que en los ejemplos típicos de transición uno no pierde contacto cognitivo con su entorno anterior y, siendo esto así, ese entorno debería ser relevante a la hora de determinar qué conceptos tiene, también después de la transición.

Digamos algo sobre auto-conocimiento. Hemos visto que los argumentos incompatibilistas suelen basarse en un modelo observacional del auto-conocimiento, y hemos afirmado que este modelo es falso (no somos muy originales en esto: la mayoría de las respuestas compatibilistas atacan este modelo observacional del auto-conocimiento). Concretamente, hemos visto que los argumentos incompatibilistas suelen asumir que el auto-conocimiento comparte con el conocimiento perceptivo, por

un lado, principios como (Discp) y (RA) y, por el otro, una noción internista de la evidencia. Hemos defendido que estas dos tesis no son "generalmente verdaderas" (esto es: no ponen condiciones a toda instancia de conocimiento), y que tampoco lo son en el caso concreto del auto-conocimiento.

Así, hemos sugerido que el conocimiento de los contenidos de nuestros estados mentales se asemeja a nuestro conocimiento de proposiciones como que estoy aquí ahora. Como dice Burge, este conocimiento de contenidos no es falible, no hay lugar para errores "brutos", y quizás deberíamos asumir que cuando uno no sabe que tiene la creencia de que *p*, eso es así, no porque no conoce el contenido de que *p*, sino porque no sabe que guarda una relación de creencia hacia ese contenido (hablando un tanto toscamente, cuando hay fallos de este tipo en el auto-conocimiento, eso es así, no porque "desconocemos contenidos", sino porque "desconocemos actitudes").

Recordemos también que en la segunda parte hemos esbozado lo que hemos llamado una *(proto)teoría de la memoria*. Habíamos introducido el ejemplo de una transición lenta que presentaba a un sujeto que intentaba recordar, en distintas situaciones, pensamientos que había tenido en entornos anteriores; hemos hecho explícito cuáles eran las predicciones a las que queríamos llegar, y hemos descrito las líneas generales que tendría que respetar una teoría de la memoria para poder llegar a esas predicciones. Por supuesto, el externista no está comprometido a asumir algo parecido a esta (proto)teoría, pero sí lo está aquel externista que comparta nuestras intuiciones sobre qué puede recordar alguien como Sally y qué no en qué situaciones.

Hemos comenzado diferenciando entre memoria episódica y semántica: la primera nos proporciona información y conocimiento sobre nuestras vivencias pasadas, la segunda nos "trae a la mente" las creencias y el conocimiento que hemos adquirido anteriormente. Primero, hemos mantenido que ambas funciones son preservativas: no hay motivos para pensar que los recuerdos de la víctima de una transición lenta cambian de contenido; pero, lo que quizás es más interesante, hemos defendido que estas dos memorias muestran diferencias de carácter epistémico. Nuestros recuerdos episódicos se basan en una representación mnemónica, causada por la vivencia en cuestión que representa. Cuando uno recuerda episódicamente que *p*, la representación mnemónica de aquella vivencia que recuerda forma la evidencia que tiene para recordar y saber que

p, de lo que se sigue que cuando uno recuerda episódicamente que *p*, hay un elemento de naturaleza mnemónica en su justificación. Por contra, hemos dicho que la memoria semántica simplemente preserva la justificación original, de lo que se sigue que cuando uno recuerda semánticamente que *p*, no hay ningún factor de naturaleza mnemónica en su justificación.

Terminemos este trabajo recordando que en la tercera parte afirmamos que el externismo semántico tiene una consecuencia, que a algunos les puede parecer incómoda, pero no así a nosotros. Según dijimos, se sigue, una vez aceptamos una semántica externista sobre los estados mentales, y seguimos con nuestras prácticas habituales de racionalización y explicación de conducta, que algunos factores externos (como la veracidad o no de sus creencias de identidad) pueden ser relevantes a la hora de determinar qué premisas que contiene la inferencia que sigue un sujeto. Por ejemplo, S y S* pueden ser internamente indistinguibles, teniendo S una creencia verdadera de identidad que expresaría profiriendo el enunciado '*a*=*a*', y S* una creencia falsa de identidad que expresaría profiriendo el mismo enunciado '*a*=*a*'. Pero hemos defendido que, si esto es así, entonces hay una inferencia que daría a conocer S mediante la alocución L que no contiene como premisa la creencia de identidad que expresaría S mediante la proferencia de '*a*=*a*', y que hay otra inferencia que daría a conocer S* mediante la misma alocución L que sí contiene como premisa la creencia de identidad falsa que expresaría S* mediante la proferencia de '*a*=*a*'. Esto es, las propiedades internas de un sujeto no bastan para determinar qué premisas tienen las inferencias que sigue.

(Bibliografía)

- Audi, Robert (1998): *Epistemology: a contemporary introduction to the theory of knowledge*, Rutledge, London

- Bernecker, Sven (1998): "Self-Knowledge and Closure", en Ludlow y Martin (1998)
  - (2004) "Externalism and Memory", *Philosophical and Phenomenological Research 69*
  - (2008) *The Metaphysics of Memory*, Springer

- Boghossian, Paul (1989): "Content and Self-Knowledge", *Philosophical Topics 17* (citas desde Ludlow y Martin (1998))
  - (1992a) "Externalism and Inference", *Philosophical Issues Vol.2, Rationality in Epistemology.*
  - (1992b) "Reply to Schiffer", *Philosophical Issues Vol. 2, Rationality in Epistemology*.
  - (1994) "The Transparency of Mental Content", *Philosophical Perspectives Vol. 8, Logic and Language*

- Brown, Jessica (1998): "Natural Kind Terms and Recognitional Capacities", *Mind 107*
  - (2000) "Critical Reasoning, Understanding, and Self-Knowledge", *Philosohpy and Phenomenological Research 61*
  - (2004) *Anti-individualism and Knowledge*, MIT Press, Cambridge

- Brueckner, Anthony (1997): "Externalism and Memory", *Pacific Philosophical Quarterly 78* (citas desde Ludlow y Martin (1998)).

- Burge, Tyler (1977): "Belief *De Re*", *Journal of Philosophy 74*
  - (1979) "Individualism and the Mental", *Midwest Studies in Philosophy 4* (citas desde Burge (2007a))
  - (1982) "Other Bodies" en Woodfield (ed.): *Thought and Content*, Oxford University Press. (citas desde Burge (2007a)).
  - (1988) "Individualism and Self-Knowledge", *The Journal of Philosophy 85*
  - (1989) "Wherein is Language Social?", en George, A. (ed.): *Reflecionts on Chomsky*, Blackwell, Oxford (citas desde Burge (2007a)
  - (1993a) "Content Preservation", *The Philosophical Review 4*
  - (1993b) "Concepts, Definitions, and Meaning", *Metaphilosophy 24* (citas desde Burge (2007a))
  - (1996) "Our entitlement to Self-Knowledge", *Proceedings of the Aristotelian Society 96* (citas desde Ludlow y Martin (1998))
  - (1998) "Memory and Self-Knowledge", en Ludlow y Martin (1998)
  - (2007a) *Foundations of Mind*, Oxford University Press
  - (2007b) "Postscript to "Individualism and the Mental", en Burge (2007a)

- Campbell, John (1987): "Is Sense Transparent?", *Proceedings of the Aristotelian Society 88*

- Collins, John M. (2008): "Content Externalism and Brute Logical Error"*, Canadian Journal of Philosophy 38*

- Crimmins, Mark y John Perry (1989): "The Prince and the Phoneboot", *The Journal of Philosophy 86*

- Davidson, Donald (1987): "Knowing One's own Mind", *Proceedings of the American Philosophical Association* (citas desde Ludlow y Martin (1998))

- Donnellan, Keith (1977): "The Contingent A Priori and Rigid Designation", *Midwest Studies in Philosophy 2*

- Dummett, Michael (1973): *Frege: Philosophy of Language*, Duckworth, Londres.
  - (1978) *Truth and Other Enigmas*, Duckworth, Londres.

- Evans, Gareth (1982): *The Varieties of Reference*, Oxford University Press, Oxford

- Falvey, Kevin (2003): "Memory and Knowledge of Content", en Nuccetelli (2003)

- Falvey, Kevin y Joseph Owens (1994): "Externalism, Self-Knowledge, and Skepticism", *The Philosophical Review 103*

- Faria, Paulo (2009): "Unsafe Reasoning: a Survey", *Dois Pontos 5*

- Fernández, Jordi (2006): "Intentionality of Memory", *Australasian Journal of Philosophy*.

- Fodor, Jerry (1998): *Concepts: Where Cognitive Science went wrong*, Oxford University Press.

- Frege, Gottlob (1892): "Über Sinn und Bedeutung", traducción al castellano "Sobre Sentido y Referencia" *Escritos Filosóficos*, Cátedra, Barcelona, 1996.
  - (1914) "Logic in Mathematics", en *Posthumous Writings*, Oxford, Blackwell, 1979

- Gaiman, Neil (2002): *Coraline*, HarperCollins Publishers

- Gertler, Brie (ed.) (2003): *Privileged Access: Philosophical Accounts of Self-Knowledge*, Ashgate

- Gibbons John (1996): "Externalism and Knowledge of Content", *The Philosophical Review 105*

- Goldberg, Sanford (1999): "The Relevance of Discriminatory Knowledge of Content", *Pacific Philosophical Quarterly 80*
  - (2000) "Externalism and Authoritative Knowledge of Content: a New Incompatibilist Strategy", *Philosophical Studies 100*
  - (2003) "Anti-individualism, Conceptual Omniscience and Skepticism", *Philosophical Studies 116*
  - (2006) "Brown on Self-Knowledge and Discriminability", *Pacific Philosophical Quarterly 87*
  - (2007a) (ed.) *Internalism and Externalism in Semantics and Epistemology*, Oxford University Press, Oxford

- (2007b) “Semantic Externalism and Epistemic Illusions”, en Goldberg (2007a)
- (2008) “Must Differences in Cognitive Value be Transparent?”, en *Erkenntnis 69*

- Goldman, Alvin ((1976): “Discrimination and Perceptual Knowledge”, *The Journal of Philosophy 73*
  - (1986) *Epistemology and Cognition*, Harvard University Press, Cambridege

- Heal, Jane (1998): “Externalism and Memory (II)”, *Proceedings of the Aristotelian Society 72*

- Heil, John (1988): “Privileged Access”, *Mind 97*

- Jeshion, Robin (2010): “Singular Thought: Acquaintance, Semantic Instrumentalism, and Cognitive Elasticity”, en Jeshion (ed.): *New Essays on Singular Thought*, Oxford University Press, 2010

- Kaplan, David (1970): “Dthat”, en *Syntax and Semantics 9* (notas desde Martinich (2008))
  - (1989) “Demonstratives” en Almog, Wettstein y Perry (eds.): *Themes from Kaplan*, Oxford University Press

- Kraay, Klaas J. (2002): “Externalism, Memory, and Self-Knowledge”, *Erkenntnis 56*

- Kripke, Saul (1979): “A Puzzle about Belief”, en Avishai Margalit (ed.): *Meaning and Use*, Reidel Publishing Company.
  - (1980) *Naming and Necessity*, Harvard University Press, Cambridge

- Ludlow, Peter (1995a): “Externalism, Self-Knowledge, and the Prevalence of Slow Switching”, *Analysis 55* (citas desde Ludlow y Martin (1998)
  - (1995b) “Social Externalism, Self-Knowledge and Memory”, *Analysis 55* (citas desde Ludlow y Martin (1998))
  - (1996) “Externalism and Memory: A Problem?”, *Acta Analytica 14*, (citas desde Ludlow y Martin (1998))
  - (1997) “On the Relevance of Slow Switching”, *Analysis 57* (citas desde Ludlow y Martin (1998))
  - (1999) “First Person Authority and Memory”, en Mario de Caro (ed.): *Interpretation and Causes: New Perspectives on Donald Davidson’s Philosophy*, Kluwer.

- Ludlow, Peter y Norah Martin (ed) (1998): *Externalism and Self-Knowledge*, CSLI Publications

- Martin, C. B. y Max Deutscher (1966): “Remembering”, *The Philosophical Review 75*

- Martinich, A. P. (2008): *The Philosophy of Language*, OUP, Oxford

- McDonald, Cynthia, Barry Smith y Crispin Wright (eds.) (1998): *Knowing Our Own Minds*, Oxford University Press, Oxford

- McGinn, Colin (1984): "The Concept of Knowledge", *Midwest Studies in Philosophy 9*
  - (1989) *Mental Content*, Blackwell, Oxford.

- McKinsey, Michael (1991): "Anti-individualism and Privileged Access", *Analysis 51*

- McLaughlin, Brian y Michael Tye (1998): "Is Content-Externalism compatible with Privileged Access?", *The Philosophical Review 107*

- Nagel, Thomas (1976): "Moral Luck", *Proceedings of the Aristotelian Society 50*

- Nuccetelli, Susana (ed) (2003): *New Essays on Semantic Externalism and Self-Knowledge*, MIT Press

- Owens, Joseph (1989): "Contradictory Beliefs and Cognitive Access", en *Midwest Studies Vol. 14 Contemporary Perspectives in Philosophy of Language II*.
  - (1990) "Cognitive Access and Semantic Puzzles", en Anthony Anderson y Joseph Owens (eds.): *Propositional Attitudes: The Role of Content in Logic, Language and Mind*. CSLI Press, Palo Alto.

- Pérez Otero, Manuel (2009): "Conocimiento, discriminabilidad, y acceso al contenido representacional", en Alcolea, Iranzo, Sánchez y Valor (eds.): *Actas del VI Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia*, Universitat de Valencia, Valencia.

- Perry, John (1977): "Frege on Demonstratives", *Philosophical Review 86*
  - (1979) "The Problem of the Essential Indexical", *Nous 13*
  - (1988) "Cognitive Significance and New Theories of Reference"
  - (2001) *Reference and Reflexivity*, CSLI Publications.

- Putnam, Hilary (1973): "Meaning and Reference", en *Journal of Philosophy 70* (citas desde Martinich (2008)).
  - (1975) "The Meaning of 'Meaning'", en Hilary Putnam: *Mind, Language and Reality*, *Philosophical Papers II*, Cambridge University Press, Cambridge

- Russell, Bertrand (1910): "Knowledge by Acquaintance and Knowledge by Description", *Proceedings of the Aristotelian Society 11*

- Salmon, Nathan (1986): *Frege's Puzzle*, MIT Press
  - (1989) "Illogical Belief", *Philosophical Perspectives 3, Philosophy of Mind and Action Theory*

- Schiffer, Stephen (1992): "Boghossian on Externalism and Inference", *Philosophical Issues* 2

- Searle, John (1982): *Intentionality*, Cambridge University Press, Cambridge

- Shoemaker, Sydney (1994): "Self-Knowledge and 'Inner Sense': Lectures I, II and III", *Philosophy and Phenomenological Research 54*

- Soames, Scott (1988): "Direct Reference, Propositional Attitudes, and Semantic Content", *Philosophical Topics 15*
    - (2002) *Beyond Rigidity*, Oxford University Press, Oxford

- Sorensen, Roy A. (1998): "Logical Luck", *The Philosophical Quarterly 48*

- Sosa, David (2007): "The Inference that Leaves Something to Chance", en Goldberg (2007a)

- Stalnaker, Robert (2008); *Our Knowledge of the Internal World*, Oxford University Press, Oxford

- Tulving, Endel (1972): "Episodic Memory and Semantic Memory", en Endel Tulving y Wayne Donaldson (eds.): *Organization of Memory*, Academic Press, Londres
    - (1990) "Episodic Memory", en Michael Eysenck (ed.): *The Blackwell Dictionary of Cognitive Psychology*, Blackwell, Cambridge
    - (2001) "Episodic Memory and Common Sense: How far apart?", *Philosophical Transactions of the Royal Society B*
    - (2002) "Episodic Memory: From Brain to Mind", *Revue Neurologique 160*

- Tye, Michael (1998): "Externalism and Memory (I)", *Proceedings of the Aristotelian Society 72*

- Warfield, Ted A. (1992): "Privileged Self-Knowledge and Externalism are Compatible", *Análisis 52* (citas desde Ludlow y Martin (1998))
    - (1997) "Externalism, Privileged Self-Knowledge, and the Irrelevance of Slow Switching", *Análisis 57* (citas desde Ludlow y Martin (1998))

- Wikforss, Åsa (2001): "Social Externalism and Conceptual Errors", *Philosophical Quarterly 51*
    - (2004) "Externalism and Incomplete Understanding", *Philosophical Quarterly 54*

- Williams, Bernard (1976): "Moral Luck", *Proceedings of the Aristotelian Society 50*

- Williamson, Timothy (1990): *Identity and Discrimination*, Basil Blackwell , Oxford
    - (1995) "Is Knowing a State of Mind?", en *Mind 104*
    - (2000) *Knowledge and its Limits*, Oxford University Press, Oxford.