

LA PRUEBA DE PUTNAM CONTRA EL ESCEPTICISMO RADICAL: DOS REINTERPRETACIONES BASADAS EN EL AUTOCONOCIMIENTO

MANUEL PÉREZ OTERO

Departamento de Lógica, Historia y Filosofía de la Ciencia

Facultad de Filosofía

Universidad de Barcelona/LOGOS

(Research Group in Logic, Language and Cognition)

perez.otero@ub.edu

RESUMEN: Presento y defiendo dos reinterpretaciones de la prueba ideada por Putnam para demostrar que no somos cerebros en una cubeta. Ambas resaltan explícitamente el papel desempeñado por el autoconocimiento de nuestros propios pensamientos y por el externismo sobre el contenido. La primera asume que el externismo implica que un cerebro en una cubeta no puede pensar el contenido proposicional relevante (constituido por conceptos acerca de cerebros y cubetas). La otra versión invoca una tesis externista más débil, conforme a la cual un sujeto corpóreo y un cerebro en una cubeta no pueden compartir el contenido proposicional relevante.

PALABRAS CLAVE: cerebros en cubetas, Falvey y Owens, externismo sobre el contenido, *petitio principii*, justificación epistémica

SUMMARY: I present and defend two reinterpretations of Putnam's proof that we are not brains in a vat. Both of them make explicit the role played in it by self-knowledge of our own thoughts and the role played by content externalism. The first one takes externalism as implying that a brain in a vat cannot have the relevant content (involving concepts about brains and vats). The other version only extracts from externalism a weaker thesis that a bodily subject and a brain in a vat cannot share the relevant content.

KEY WORDS: brains in a vat, Falvey and Owens, content externalism, *petitio principii*, epistemic justification

1. *Introducción*

En el capítulo inicial de su libro *Reason, Truth and History*, Hilary Putnam elaboró un argumento mediante el cual pretendía demostrar que no somos cerebros en una cubeta. Presentaré dos reinterpretaciones de su argumento. En ambas se resalta explícitamente el papel clave desempeñado por el autoconocimiento de los propios pensamientos. Una de ellas (cercana a la interpretación de Wright 2000) se corresponde mejor con la literalidad del texto de Putnam y, presumiblemente, con sus intenciones. Puede verse como una interpretación que aspira a reconstruir el argumento original. La otra no está tan

apegada al texto de Putnam. Pero también resulta bastante natural entresacarla a partir de lo que Putnam propone, y no es implausible suponer que también haya tenido en mente algo así.

Defenderé las dos versiones. Con respecto a la primera, rechazaré una objeción (basada en una crítica similar formulada por Kevin Falvey y Joseph Owens contra una versión más abstracta del argumento) que acusa al argumento de incurrir en petición de principio. En relación con la segunda reinterpretación, ofreceré justificación, en positivo, de por qué es un buen argumento, clarificando las motivaciones principales del escepticismo que dicho argumento (así como también la primera reinterpretación y el argumento original de Putnam) trata de combatir.

2. *Autoconocimiento del contenido*

Putnam ofrece varias caracterizaciones, no enteramente coincidentes, de la situación en la que están los cerebros en una cubeta de los que trata su experimento mental. Pero las diferencias no son muy importantes. Casi todas esas caracterizaciones comparten los siguientes rasgos. Pensemos en cerebros, desprovistos de cuerpos, que experimentarían sensaciones —provocadas artificialmente— subjetivamente indistinguibles de las sensaciones que experimenta un sujeto normal, corpóreo. Esos cerebros están (han estado siempre, durante toda su existencia) en una cubeta con nutrientes que los mantienen vivos, y sus terminaciones nerviosas se conectan con un ordenador, cuyo complejo programa permite que el “sujeto”, el cerebro, tenga aparentes percepciones de un mundo externo, cualitativamente equivalentes a las que tiene cualquiera de nosotros, quienes —suponemos— somos personas corpóreas. En una primera aproximación, Putnam imagina que el ordenador es manejado por un científico maligno.¹ Pero luego precisa el escenario de otra forma, estipulando que ningún ser inteligente ha diseñado ni maneja el ordenador, que se ha originado casualmente o ha existido siempre.² Asumiremos que el escenario es este último.³

¹ *Cfr.* Putnam 1981, pp. 5–6.

² *Cfr.* Putnam 1981, p. 12.

³ Cabe suponer que con ello Putnam quiere facilitarse las cosas, evitando el riesgo de que un objetor rechace su tesis fundamental (que el cerebro en la cubeta no puede tener pensamientos acerca de cerebros y cubetas) alegando que el cerebro adquiere deferencialmente conceptos como los de cerebro y cubeta a través del científico, que posee de forma estándar tales conceptos. De todos modos, creo que Putnam consideraría que su argumento también vale contra el otro escenario (y por ello su primera descripción de la situación no excluye al científico), porque las

La tesis crucial en la argumentación desarrollada por Putnam para concluir que no estamos en esa situación es que si fuéramos cerebros en una cubeta, *no podríamos decir o pensar* que lo somos (1981, p. 7). Es metafísicamente posible (admite Putnam) que haya cerebros en cubetas. Pero tenemos recursos conceptuales suficientes para descartar legítimamente que seamos uno de ellos; podemos probar que no lo somos. Utilizando vocabulario algo más técnico: aunque es metafísicamente posible que haya cerebros en cubetas (quizá incluso —esta cuestión no la trata Putnam— es metafísicamente posible que *nosotros* seamos cerebros en cubetas), no es epistémicamente posible (no es compatible con todo lo que sabemos) que seamos cerebros en cubetas. Ese contraste está presente en otros casos más triviales: aunque es metafísicamente posible que yo nunca haya existido (hay mundos posibles en que jamás llego a existir), tal cosa no es epistémicamente posible (sé que no estoy en uno de esos mundos posibles).

Vale la pena poner de relieve que la estrategia argumentativa de Putnam es de tipo *trascendental*. Se investigan e invocan las *precondiciones* requeridas para que *pueda* haber representaciones (las *condiciones de posibilidad* de que haya representaciones).⁴ Para que esa estrategia tenga alguna viabilidad es crucial que la hipótesis escéptica contemplada sea una hipótesis especialmente radical, tal y como se ha descrito en el penúltimo párrafo: que los cerebros hayan estado *siempre* en esa situación.⁵

Un primer paso hacia la clarificación del argumento (que emprenderemos con más detalle en secciones posteriores) es constatar que dicha tesis crucial no permite alcanzar la conclusión prometida. La exposición de Putnam permite atribuirle una versión fuerte de esa tesis, no restringida a nosotros, o a uno mismo. En general (es decir, para cualquier x): si x fuera un cerebro en una cubeta, x no podría pensar que lo es, o que no lo es. La proposición que se pretende demostrar es que *yo no soy un cerebro en una cubeta*. (Naturalmente, los diferentes sujetos que consideren el argumento estarán considerando

conexiones causales con los objetos del entorno en el caso del científico y en el caso del cerebro en la cubeta son tan diferentes que la teoría externista correcta no podría aceptar que hubiera coincidencia en contenido representacional. (El comentario de un árbitro anónimo de *Crítica* me ha hecho ver la conveniencia de hacer referencia a esta cuestión.)

⁴ Cfr. Putnam 1981, p. 16.

⁵ En Pérez Otero 2011, sec. 4, destaco ese punto y expongo por qué el argumento resultaría inocuo contra la hipótesis de que somos cerebros recientemente “encubetados”. Cfr. también Quesada 1998, pp. 118–120, y Wright 1994, pp. 234 y ss.; y 2000, nota 22.

proposiciones diferentes, que conciernen distintivamente a cada uno de ellos.) Para completar el razonamiento conviene formular expresamente una premisa implícita que Putnam usa: *yo puedo pensar que soy (o que no soy) un cerebro en una cubeta*, puesto que *estoy considerando ahora mismo ese pensamiento*. Así pues, una versión simple de su argumento es ésta:

PUTNAM-1:

P_1 : *Estoy considerando el pensamiento de que no soy un cerebro en una cubeta.*

P_n : *Si x fuera un cerebro en una cubeta, entonces x no podría estar considerando el pensamiento de que no es un cerebro en una cubeta.*

C : *(Por lo tanto) No soy un cerebro en una cubeta.*

Es un argumento válido. La cuestión fundamental es cómo se justifican sus premisas, especialmente la segunda, P_n . Las dos versiones que después presentaré y defenderé conllevan dos formas diferentes de proporcionar justificación apropiada para P_n . De momento, digamos algo sobre P_1 .

La primera premisa, P_1 , está respaldada por los supuestos usuales sobre la autoridad especial de la primera persona en relación con los contenidos de los propios pensamientos. Un sujeto que se plantea la cuestión de si es o no un cerebro en una cubeta tiene conocimiento de lo que P_1 establece. Pueden ofrecerse diversas explicaciones teóricas de dicho autoconocimiento, pero no tenemos necesidad de seleccionar entre ellas (y discutir sus ventajas y desventajas respectivas nos apartaría demasiado de nuestro tema).

No obstante, conviene constatar una importante limitación del argumento PUTNAM-1. Como veremos en seguida, la justificación que sustenta a P_n depende (en parte) de consideraciones típicamente *externistas* sobre la individuación del contenido representacional. Así pues, las dos premisas de PUTNAM-1 sólo serán aceptables para quienes estén dispuestos a admitir una posición *compatibilista* sobre el presunto conflicto entre ambas tesis (externismo y autoconocimiento).⁶

⁶He defendido ese compatibilismo en diversos trabajos: Pérez Otero 2004a, 2004b, 2009b, 2010b. Estoy en deuda con Tobies Grimaltos y Carlos Moya por ayudarme a ver la conveniencia de explicitar esa restricción en PUTNAM-1.

Sin embargo, contrariamente a lo que algunos filósofos simpatizantes del incompatibilismo pudieran sostener, para que PUTNAM-1 permita establecer lo prometido (la proposición literalmente expresada por su conclusión) no se requiere que P_1 exprese algún tipo de autoconocimiento especialmente *sustantivo* y poco trivial (cuanto más sustantivo sea, más riesgo habrá de conflicto real con el externismo). P_1 , interpretada literalmente (por poco sustantiva que sea), ya permite —en conjunción con P_n — derivar C. Si un filósofo pretende que lo que se obtiene así es muy poco y por tanto resulta insatisfactorio, es ese filósofo quien tiene la carga de la prueba a la hora de caracterizar un sentido más fuerte del significado de C (o del significado de C y/o del significado de P_1) y justificar que es ese sentido más fuerte el que estaba en juego.⁷

Por otra parte, independientemente del recurso al autoconocimiento, podemos invocar otras consideraciones para defender la legitimidad de incluir P_1 entre las premisas. Estas otras consideraciones se vinculan con el hecho de que el razonamiento de Putnam tiene las características de un argumento trascendental. Es sensato concebir el argumento como la refutación de una posición escéptica que niega (o pone en duda) que tengamos conocimiento o justificación de C. Al demostrar C (por obtener C como conclusión de un argumento deductiva o inductivamente válido), es plausible pensar que tenemos conocimiento o justificación de C, contrariamente a la pretensión de dicho escéptico hipotético. Pero la pretensión de ese escéptico quedaría también legítimamente rechazada si resultara que es ininteligible. Un requisito de inteligibilidad de su pretensión es precisamente la verdad de P_1 . En otras palabras, se requiere P_1 para asumir el planteamiento mismo de la hipótesis del escenario escéptico (el escenario cuya posibilidad metafísica invalidaría —según el escéptico— nuestro conocimiento o nuestra justificación de que no nos hayamos en él).

3. *Lo que no podrían representar los cerebros en una cubeta*

Pasemos a evaluar P_n . En un pasaje de la penúltima sección, el propio Putnam hace explícitas dos de las premisas empleadas para sustentar el argumento (que ha presentado en las páginas previas). Esas dos premisas contribuyen a respaldar P_n . Quedan resumidas (según indicaré en seguida) en la premisa P_2 , de la siguiente formulación:

⁷ Las objeciones de María José Frápolli al argumento en su versión PUTNAM-AC (tal y como lo presento en la próxima sección) me han llevado a incluir este párrafo aclaratorio.

PUTNAM-AC:

P_1 : *Estoy considerando el pensamiento de que no soy un cerebro en una cubeta.*

P_2 : *Es verdadera una teoría mínimamente causal y externista sobre el significado.*

P_n : *Si x fuera un cerebro en una cubeta, entonces x no podría estar considerando el pensamiento de que no es un cerebro en una cubeta.*

C : *(Por lo tanto) No soy un cerebro en una cubeta.*

Las letras “AC” en esa denominación están por *autoconocimiento*. Las incluyo porque esa versión del argumento pone de manifiesto (más claramente que en las formulaciones de Putnam) el papel que desempeña el autoconocimiento del contenido de nuestros pensamientos. Obviamente, eso ocurría ya con PUTNAM-1, que cabe considerar como un paso previo en la dirección de PUTNAM-AC. (Pero también PUTNAM-AC es todavía una versión muy abstracta del argumento. No coincide con ninguna de las dos reconstrucciones que quiero presentar, las cuales son especificaciones de PUTNAM-AC.) Lo que aporta distintivamente PUTNAM-AC (en relación con PUTNAM-1) no es la compleción de un entimema. Ya he señalado que PUTNAM-1 es válido; por lo tanto, P_2 no es una premisa que deba añadirse para corregirlo. El papel de P_2 es esclarecer los pasos que conducen hasta P_n .

Las dos premisas a las que se refiere Putnam son:

- (a) *las teorías mágicas de la referencia son erróneas;*
- (b) *es imposible referirse a ciertos tipos de cosas, por ejemplo a árboles, sin haber tenido interacción causal con ellas o, en su caso, con otras cosas en cuyos términos pueden ellas describirse.⁸*

En la tesis (a) se emplea, de forma sesgada, el calificativo de *mágico*, para descartar cierto tipo de teorías. Si eliminamos esos elementos retóricos y tenemos en cuenta el texto de Putnam, (a) establece aproximadamente:

⁸ Cfr. Putnam 1981, p. 16.

(a*) *El contenido semántico de una presunta representación (en particular, aquello a lo que refiere) no queda determinado por propiedades intrínsecas que comparte con otra presunta representación cualitativamente indistinguible.*

La tesis (a*) es una tesis negativa, sobre qué propiedades de una presunta representación no son suficientes para determinar su contenido. La tesis (b) es una tesis positiva, que identifica ciertos factores como necesarios para la determinación del contenido. El externismo semántico incorpora precisamente intuiciones en favor de ambos tipos de tesis. Por eso me parece apropiado resumir (a*) y (b) mediante la premisa P₂: *Es verdadera una teoría mínimamente causal y externista sobre el significado.*

En favor de la tesis negativa, (a*), Putnam ponía de manifiesto intuiciones diversas (de inspiración wittgensteiniana) contrarias a una teoría *imaginística* del significado y la referencia (teoría que Putnam clasifica como *mágica*). Ése es el tipo de teoría que (a*) descarta. Se realzan las intuiciones en su contra al considerar casos de “imágenes” indistinguibles que difieren en lo que representan, especialmente cuando una de ellas no tiene contenido representacional alguno: por ejemplo, una caricatura de Churchill trazada intencionalmente y una figura indiscernible de la anterior trazada aleatoriamente por una hormiga al arrastrarse por la arena.

La tesis positiva (b) parece bastante natural cuando uno se pregunta qué factores han de resultar relevantes (en la determinación del contenido), como alternativa a esos descartados por (a*). La intuición sería que la alternativa a esa teoría *imaginística* debe imponer ciertos requisitos mínimamente *externistas* entre una representación y su contenido, estableciendo que haya algún tipo de conexión *histórico-causal*. Es un requisito externista porque las relaciones histórico-causales relevantes no dejarían un rastro en los aspectos intrínsecos cualitativos de la representación, los únicos accesibles a un sujeto que examinase por introspección la representación si es que ésta es interna. (Putnam 1973, 1975 y Burge 1978 eran los textos fundacionales de tales posiciones externistas o antiindividualistas.)

Entre los mayores atractivos de la argumentación de Putnam se cuenta su defensa de P_n basándose en P₂ (y empleando también P₁, según hemos de clarificar). Términos generales como “tigre”, “oro” o, también, “cerebro” y “cubeta” tienen típicamente un determinado significado como consecuencia (al menos) de conexiones histórico-causales apropiadas entre los usos de tales términos y sus referentes; o entre los usos de tales términos y otros usuarios del lenguaje, que

han empleado previamente tales símbolos. Lo mismo cabe decir si hablamos de símbolos internos correlativos de tales expresiones; los símbolos que se activarían al tener algún pensamiento constituido por los conceptos respectivos. Para que un sujeto se represente que no es un cerebro en una cubeta (ya sea con una oración perteneciente a un lenguaje público, o bien meramente al creer dicho contenido) debe usar símbolos (externos o internos) apropiadamente conectados con cerebros y con cubetas (o apropiadamente conectados con otros usuarios de tales símbolos). Sólo entonces el sujeto puede poseer los conceptos de *cerebro* y *cubeta*. Los presuntos símbolos internos que manejara un cerebro en una cubeta no cumplirían esos requisitos. Tales supuestos símbolos no podrían representar, por ejemplo, cerebros y cubetas. Ese “individuo”, el cerebro en una cubeta, no podría albergar —por lo tanto— el pensamiento de que no es un cerebro en una cubeta.

El párrafo anterior condensa la transición desde P_2 (en conjunción con P_1) hasta P_n . La veremos con más detalle en las secciones 4 y 6, donde presento dos versiones diferentes de PUTNAM-AC, vinculadas respectivamente a dos interpretaciones diferentes de P_2 , una más arriesgada y otra menos comprometida. Conforme a la interpretación más arriesgada de P_2 (que probablemente se ajusta mejor a las intenciones de Putnam), esta tesis implicaría que ninguna presunta representación interna que manejara (ahora, al considerar —supuestamente— un presunto pensamiento internamente indistinguible del que considera un sujeto corpóreo) un cerebro en una cubeta podría hacer referencia a (representar) cerebros, pues carecería de la conexión causal apropiada. Conforme a la otra interpretación, menos comprometida, P_2 implicaría sólo que la conexión causal apropiada no podrían tenerla “simultáneamente” (al albergar presuntos pensamientos internamente indistinguibles) un sujeto corpóreo y un cerebro en una cubeta; implicaría, pues, que no podrían *ambos* hacer referencia a cerebros cuando aparentemente ambos están pensando que no son cerebros en una cubeta.

4. *La versión basada en la ausencia del contenido apropiado-pretendido*

Es deseable precisar el argumento PUTNAM-AC en varios aspectos. Principalmente, convendría aclarar la justificación esbozada de P_n . Interviene, es obvio, P_2 . Pero he indicado que interviene también P_1 , lo cual no resulta tan obvio. Así, un punto importante es precisar cómo P_1 contribuye a sustentar P_n . Ofrezco en esta sección una

versión de PUTNAM-AC conforme a tales *desiderata* (la otra versión se presenta en la sección 6).

Esta interpretación invoca la lectura más arriesgada (y más natural) de las consecuencias que implica P_2 (mencionada en la sección anterior). (Al formularla y en la discusión posterior, uso el término “CEREBRO” siguiendo la convención de escribir en mayúsculas una expresión para referirme al concepto expresado por esa expresión.) Llamaré VERSIÓN-A a esta reconstrucción.

VERSIÓN-A:

P_1 : *Estoy considerando el pensamiento de que no soy un cerebro en una cubeta.*

P_1^* : *La extensión del concepto CEREBRO (constituyente de ese contenido) que uso es la clase de los cerebros.*

P_2 : *Es verdadera una teoría mínimamente causal y externista sobre el significado.*

P_2^* : *Si x fuera un cerebro en una cubeta, entonces x no podría usar (ahora) un concepto cuya extensión es la clase de los cerebros.*

P_n : *Si x fuera un cerebro en una cubeta, entonces x no podría estar considerando el pensamiento de que no es un cerebro en una cubeta.*

C : *(Por lo tanto) No soy un cerebro en una cubeta.*

En la VERSIÓN-A, P_1^* y P_2^* son suficientes para derivar C . Nuevamente, las otras premisas no están porque sean imprescindibles, sino para facilitar la comprensión global del razonamiento, incluyendo su relación con la versión más general, PUTNAM-AC.

Para constatar cómo se justifica P_1^* , es pertinente invocar, en primer lugar, P_1 . El autoconocimiento de nuestros pensamientos nos permite conocer —entre otras cosas— que dichos pensamientos están constituidos por tales y cuales constituyentes. Aplicándolo a este caso, es razonable derivar de P_1 algo como: *Estoy considerando un pensamiento que tiene entre sus constituyentes el concepto CEREBRO.* Esa proposición es una parte de lo aseverado por P_1^* . Otra parte

del contenido de P_1^* también puede derivarse del autoconocimiento expresado en P_1 ; consistiría en saber algo sobre el papel lógico-gramatical de los conceptos constituyentes del pensamiento que estamos considerando, concretamente, saber que el concepto en cuestión (el concepto CEREBRO) es un concepto predicativo. La parte restante de P_1^* vendría dada por la siguiente proposición: *La extensión del concepto CEREBRO es la clase de los cerebros*. Eso derivaría de una especie de principio “desentrecorillador” aplicado a las representaciones internas (aquí, aplicado a la representación interna correspondiente al concepto CEREBRO), combinado con supuestos básicos sobre la noción de extensión. En general, si sabemos que estamos usando el concepto X , y que X es un concepto predicativo, sabemos también que la extensión del concepto X es la clase de los X .

La premisa P_2^* , que deriva de P_2 , es compatible con cualquiera de las dos posibilidades contempladas por Putnam con respecto a la presunta representación interna usada por un cerebro en una cubeta, subjetivamente equivalente a la representación interna usada por un sujeto corpóreo cuando éste usa el concepto CEREBRO:

(i) Esa representación no representa cerebros sino alguna otra cosa, causalmente conectada con ella. Putnam menciona varias opciones sobre cuál pudiera ser ese contenido alternativo: quizá hace referencia a impulsos electrónicos que contribuyen a ocasionar la experiencia de estar aparentemente pensando en cerebros, o a características concretas del programa de ordenador (el programa que hace que el cerebro tenga todas esas aparentes percepciones), o a algún otro tipo de entidades (cerebros-meramente-arentes-en-la-“imagen”, si es que el cerebro maneja imágenes) que, en cualquier caso, no son cerebros.⁹

(ii) Esa presunta representación no tiene contenido alguno, porque un cerebro en una cubeta no tiene representaciones, no se representa nada.¹⁰

Ninguna de las dos posibilidades es compatible con que el cerebro en la cubeta esté usando (ahora, al considerar —supuestamente— un presunto pensamiento internamente indistinguible del que considera un sujeto corpóreo) un concepto cuya extensión es la clase de los cerebros. Es decir, cualquiera de las dos posibilidades, (i) y (ii), respalda la premisa P_2^* . Ya he mencionado (en la sección 3) el tipo de justificación ofrecida por Putnam en favor de que (i) y (ii)

⁹ Cfr. Putnam 1981, pp. 14–15.

¹⁰ Está claro que Putnam no descarta esta otra posibilidad; cfr. Putnam 1981, pp. 14–15.

son las únicas posibilidades abiertas. Procede de P_2 , conforme a la interpretación comprometida de esta premisa que allí se mencionó.

Por intervenir, en P_1^* , el principio “desentrecorillador” al cual me he referido, esta VERSIÓN-A resulta cercana a la reconstrucción del argumento de Putnam que propone Wright (2000). Como he anticipado, desde un punto de vista exegético la VERSIÓN-A me parece una buena interpretación de lo que Putnam tenía en mente. La VERSIÓN-B que voy a examinar en la sección 6 no está tan apegada al texto de Putnam, aunque —como espero mostrar— es también una reinterpretación razonable y sólida.

5. *La objeción de circularidad*

Nuestra primera aproximación al argumento de Putnam, la formulación PUTNAM-1, aparece discutida por Falvey y Owens (1994, pp. 132 y ss.). Ellos rechazan el razonamiento. Alegan que depende de proposiciones empíricas que serían falsas si C fuera falso y —por ello— apoyarse en tales proposiciones para tratar de demostrar C prejuzga la cuestión, y convierte en ilegítimo el argumento. Así, por ejemplo:

A fin de establecer que las preferencias del [cerebro en la cubeta] de las palabras “cerebro” y “cubeta” difieren en referencia con respecto a nuestras preferencias de esas palabras, tenemos que apelar a la premisa de que nuestras palabras “cerebro” y “cubeta” hacen referencia a ciertos objetos físicos determinados de nuestro entorno. Pero todas éstas son proposiciones empíricas, las cuales de hecho son verdaderas sólo si no somos cerebros en cubetas. Como tal, el recurso a esta premisa es ilegítimo en el presente contexto. Precisamente prejuzga la cuestión que nos ocupa. [...] Mi conocimiento de que “agua” hace referencia a agua, o de que “cerebro” hace referencia a cerebros no me permitirá detectar diferencias entre la referencia de estas palabras cuando yo las uso y la referencia de esas palabras cuando las usa el cerebro en la cubeta. (Falvey y Owens 1994, pp. 135–136.)

La proposición que he añadido a PUTNAM-1 para clarificarlo es P_2 (*Es verdadera una teoría mínimamente causal y externista sobre el significado*), y obtener así PUTNAM-AC. Falvey y Owens aluden a otras proposiciones; pero sus declaraciones permiten asumir que pretenderían aplicar su crítica también específicamente a P_2 . La premisa crucial, P_n , se ve apoyada por una proposición empírica como P_2 (o alguna otra similar) solamente si C es verdadera. Según esta objeción, eso deslegitimaría el argumento (a PUTNAM-1 o a PUTNAM-AC) y lo haría incurrir en cierto tipo de circularidad viciosa.

Dedico esta sección a examinar y rechazar dicha crítica. La crítica —si es correcta— también se aplicaría a la VERSIÓN-A. Así, podemos analizar el asunto teniendo a la vista esa versión, más precisa, del argumento.¹¹

Falvey y Owens no ofrecen ninguna caracterización del tipo de circularidad que según ellos aqueja a PUTNAM-1 (y que, adaptando su crítica a la VERSIÓN-A, aquejaría también a esta última). Conviene constatar, en primer lugar, que el mero hecho de que cierta premisa (P_n) dependa de proposiciones (como P_2) que serían falsas si la conclusión (C) fuera falsa¹² *no puede bastar para deslegitimar un argumento*. Usando algunos otros supuestos razonables, eso permitiría clasificar como ilegítimo *cualquier* argumento válido. Pensemos momentáneamente en esta “circularidad” *obviamente benigna*, presente en todos los argumentos válidos: quien rechaza la conclusión no podría racionalmente aceptar las premisas y las proposiciones que supuestamente la apoyan. No acuso a Falvey y Owens de detectar (en PUTNAM-1) esa suerte de circularidad y tomarla erróneamente por viciosa. Muy probablemente usan implícitamente alguna concepción de prejuzgar la cuestión cercana a la segunda de las definiciones que voy a mencionar. Pero tal vez sea útil didácticamente tener presente esa “circularidad” benigna, para estar advertidos de que una acusación de *petitio* resultará inapropiada si involucra rechazar cualquier argumento que sea “circular” en ese sentido tan general.

Especificaré dos de las nociones de *petitio principii* que con más frecuencia se usan en la literatura sobre el tema. Conforme a ninguna de ellas hay buenas razones para creer que la VERSIÓN-A incurra en *petitio*.

La primera definición procede de Copi (1961). Dice así: un argumento comete *petitio* si y sólo si una de las proposiciones que aparece como premisa es idéntica a la conclusión (incluso si difieren en su

¹¹ Quedará claro en la discusión que la objeción no puede tener éxito contra PUTNAM-1 o contra PUTNAM-AC si —como pretendo mostrar— no es apropiada contra la VERSIÓN-A. Por otro lado, aunque probablemente Falvey y Owens también aplicarían la objeción a la VERSIÓN-B (de la cual hablaré en la sección 6), resulta más natural (y menos arriesgado para el objetor) dirigirla contra la VERSIÓN-A. En cualquier caso, mi contrarréplica a esa objeción serviría también para defender de ella a la VERSIÓN-B.

Los contenidos de esta sección se exponen con mayor detalle en las secciones 3–9 de Pérez Otero 2012, donde se examina la objeción en relación con PUTNAM-AC. Remito a ese artículo para una exposición y justificación más pormenorizada de los conceptos y tesis aquí propuestos.

¹² Cfr. Falvey y Owens 1994, pp. 135–136.

formulación).¹³ Claramente, la VERSIÓN-A no incurre en *petición de principio* en ese sentido. Ninguna de sus premisas expresa el mismo contenido expresado por su conclusión. Resulta todavía más obvio que tampoco PUTNAM-I o PUTNAM-AC tienen ese defecto.¹⁴

La otra definición está inspirada en Jackson (1987). No obstante, la formularé en términos considerablemente neutrales y generales, que permitirían atribuírsela también a otros varios autores que han ofrecido caracterizaciones muy similares del mismo concepto, si bien algunos de ellos no creen que el fenómeno caracterizado sea el de la *petitio* (por ejemplo, y paradigmáticamente, Pryor 2004). Así, se constatará que ésta es una noción muy extendida en los debates contemporáneos.

En la práctica argumentativa, cuando proponemos un razonamiento con premisas Q_1, \dots, Q_n y conclusión D , con frecuencia ofrecemos también (explícita o implícitamente) al destinatario del razonamiento los indicios E_1, \dots, E_n que supuestamente confieren apoyo a las premisas Q_1, \dots, Q_n , respectivamente. Alguno de esos elementos, E_k , es explícito cuando, por ejemplo, es una premisa P_j , anterior a P_k . Eso sucede con la VERSIÓN-A. El elemento ofrecido como apoyo en favor de P_2^* es (o incluye) P_2 .

Sea A un razonamiento con premisas Q_1, \dots, Q_n y conclusión D . Sean E_1, \dots, E_n los elementos que supuestamente apoyan, respectivamente, las premisas Q_1, \dots, Q_n . Ésta es la definición de *petitio* inspirada en Jackson:

PJ: El argumento A prejuzga la cuestión, o incurre en *petitio principii*, si y sólo si para cualquier sujeto sería irracional combinar estas dos actitudes proposicionales: creer No D y creer que para cada E_i y Q_i , E_i apoya Q_i (incluso aunque no fuera irracional creer No D).¹⁵

El concepto de apoyo al que se hace referencia debe entenderse conforme a las siguientes reflexiones. Aunque yo crea No D (por

¹³ Cfr. Copi 1961, p. 97.

¹⁴ En la sección 9 de Pérez Otero 2012, considero una versión más laxa de la definición de Copi, inspirándome en sugerencias de Davies 2009. Tampoco ese sentido más laxo incluye casos como la VERSIÓN-A, PUTNAM-I o PUTNAM-AC.

¹⁵ Cfr. Jackson 1987, pp. 111–112. Esta caracterización incorpora, en la frase entre paréntesis, una modificación en la definición original de Jackson (aparte de otras diferencias superficiales en la formulación). El propósito es evitar una importante objeción a Jackson presentada por Hazlett (2006, pp. 349–350). Cfr. Pérez Otero 2011, sección 7.

ejemplo, aunque crea que ningún estudiante se ha comido el chocolate), puedo racionalmente aceptar que E_i (Laura dice que el estudiante Alejandro se ha comido el chocolate) apoya Q_i (el estudiante Alejandro se ha comido el chocolate). Si creo No D (y soy racional) será porque creo que otros elementos que apoyan No D tienen mayor peso que E_i . Así, el argumento con premisa Q_i , para la que se ofrece como indicio E_i , y cuya conclusión es D (un estudiante se ha comido el chocolate) no prejuzga la cuestión porque un sujeto puede racionalmente creer que E_i apoya a Q_i incluso creyendo No D. Por el contrario, un sujeto que incluyera en su sistema de creencias la proposición de que *todo lo que dice Laura sobre Alejandro se lo inventa* no podría racionalmente creer que E_i apoya Q_i (independientemente de que creyera o no Q_i).¹⁶

Tal vez alguien piense que, aceptando esas tesis sobre cómo entender la relación de apoyo, resultará especialmente difícil encontrar argumentos que incurran en *petitio* según PJ, especialmente dada la cuantificación universal presente en PJ. No. El argumento CEBRA ilustra paradigmáticamente el concepto (una vez que asumimos que en favor de su primera premisa se ofrece implícitamente la experiencia perceptiva consistente en cierta apariencia de ver una cebra).

CEBRA:

Eso es una cebra.

Ninguna cebra es una mula pintada para que parezca una cebra.

(Por lo tanto) Eso no es una mula pintada para que parezca una cebra.

Cualquier sujeto que niegue la conclusión será irracional si acepta que dicha apariencia perceptual confiere apoyo alguno —en ese caso particular— a la primera premisa. Otro argumento que ejemplifica también con claridad el fenómeno está temáticamente emparentado

¹⁶ Esa idea de apoyo se emparenta con la noción de compromiso racional utilizada por Pryor (2004), más que con su noción de justificación. Justamente por eso esta definición no difiere sustancialmente de la caracterización que ofrece Pryor del concepto de *ineficacia dialéctica* (así es como denomina al tipo de circularidad involucrada), formulada recurriendo al concepto de compromiso racional y otros análogos. (Cfr. Pryor 2004, pp. 363 y ss.; cfr. también Pérez Otero inédito a.)

con nuestra discusión. Se trata de MOORE-BIV, concebido de forma análoga a CEBRA, asumiendo que su primera premisa recibiría sustento de la apariencia perceptual de ver nuestras manos:

MOORE-BIV:

Tengo manos.

Si tengo manos, entonces no soy un cerebro en una cubeta.

*(Por lo tanto) No soy un cerebro en una cubeta.*¹⁷

No encontramos en la VERSIÓN-A (ni en PUTNAM-1 o PUTNAM-AC) nada similar a lo que sucede en CEBRA y MOORE-BIV que nos haga sospechar que es circular en el sentido de la definición de Jackson (PJ). Las declaraciones de Falvey y Owens hacen pensar que sería P_2^* la premisa cuyo apoyo a través de P_2 (procedente, en última instancia, de las intuiciones mínimamente externistas) debe negar quien crea ser un cerebro en una cubeta. Pero eso es altamente controvertido. Pretender que la VERSIÓN-A comete *petitio* conforme a JC implica que *todo* sujeto que niegue C debe (si es racional) negar tener intuiciones externistas que confieran cierto apoyo a P_2 y, a través de P_2 , a P_2^* . ¿Qué teoría sobre el apoyo epistémico basado en intuiciones puede pretender tal cosa? Sería similar a sostener que *ningún* filósofo internista puede reconocer compartir intuiciones externistas que apoyan la teoría contraria (o viceversa). O sostener que quien crea que Alejandro no se ha comido el chocolate debe negar que la declaración de Laura apoye lo contrario. En controversias teóricas científicas, filosóficas o de otra índole es común que (al menos algunos de) quienes defienden una posición acepten que la posición contraria también tiene cierto apoyo epistémico. Es sólo que (como se ilustró con el ejemplo del chocolate) creen que su posición está cualitativa y/o cuantitativamente mejor respaldada.

La crítica de Falvey y Owens podría entenderse de forma algo distinta.¹⁸ Sería también P_2^* la premisa conflictiva, pero el problema

¹⁷ Dretske (1970) construyó argumentos como CEBRA (sin respaldarlos) para criticar el principio de clausura del conocimiento. MOORE-BIV es una variación de la prueba de Moore (1939) de la existencia del mundo externo. Sobre la conexión entre dicha prueba (y otras variantes de la misma), el concepto de *petitio* y la transmisión de la justificación, *cf.* Pérez Otero 2006, 2009a, 2011, inédito a e inédito b).

¹⁸ Un árbitro anónimo de *Crítica* me ha sugerido esta otra interpretación de la objeción de Falvey y Owens.

residiría en el presunto apoyo que esta premisa recibe de P_2 (no en el apoyo que pudiera recibir P_2 de otras consideraciones). Falvey y Owens estarían señalando que quien rechaza C puede aceptar que tenemos intuiciones externistas favorables a P_2 e incluso aceptar una teoría causal-externista sobre el significado (según establece P_2). Pero ese objetor a C pondría reparos a la última fase de la presunta justificación de P_2^* : negaría que P_2 , por sí misma, fuera suficiente para justificar P_2^* . Más concretamente, el punto crucial supuestamente contenido en P_2^* sería la tesis de que *para tener el concepto CEREBRO se debe estar causalmente relacionado (del modo apropiado) con cerebros*. Eso sólo se sabría con información empírica sobre qué tipo de concepto es el concepto CEREBRO, sobre qué tipo de cosas son las entidades en su extensión; se requeriría saber que son —como dicen Falvey y Owens en la cita anterior— *ciertos objetos físicos determinados del entorno*.¹⁹ La teoría causal-externista mencionada en P_2 , presumiblemente obtenida *a priori*, no confiere apoyo alguno a esa tesis (según esta interpretación de la objeción de Falvey y Owens). En ese sentido, la VERSIÓN-A sería una *petitio* en el sentido PJ, pues *todo* sujeto que niegue C (también incluso este sujeto ahora contemplado, que niega C aunque es externista) rechazará que P_2 pueda proporcionar apoyo a P_2^* . P_2 no podría apoyar a P_2^* porque no es suficiente para apoyar cierta tesis Q, que sería necesaria para sustentar P_2^* :

Q: El concepto CEREBRO (que uso) tiene en su extensión *ciertos objetos físicos determinados del entorno*.

Y quien negara C negaría también que los supuestos datos empíricos favorables a Q puedan realmente apoyar Q (por las mismas razones por las que negaría que los supuestos datos empíricos favorables a “hay un árbol a mi derecha” —la impresión sensorial— puedan realmente apoyar esta proposición).

Para replicar a esa objeción voy a sostener que no es cierto que el sujeto deba aceptar la tesis Q como requisito para aceptar que P_2 confiere *cierto apoyo* epistémico a P_2^* . Antes, será útil —para esclarecer las cosas— recapitular cuál es la situación epistémica del sujeto que

¹⁹ Parece natural suponer que lo que quieren decir Falvey y Owens al hablar de *ciertos objetos físicos determinados del entorno* estaría en contraposición con las alternativas sobre la referencia del símbolo (público o interno) “cerebro” que menciona Putnam; es decir, no serían *objetos físicos determinados del entorno* entidades como: características concretas de un programa de ordenador, cerebros-meramente-aparentes-en-la-“imagen”, etcétera.

estamos considerando y recordar qué está en juego exactamente tras la acusación de que el argumento, en esta VERSIÓN-A, comete *petitio* según PJ (teniendo en cuenta que la VERSIÓN-A no comete *petitio* en el sentido definido por Copi, y que Falvey y Owens no ofrecen ninguna definición alternativa de la *petitio*).

El sujeto escéptico que estamos considerando, al que voy a llamar Keanu, cree ser un cerebro en una cubeta, pero acepta el externismo de P₂.²⁰ Dicho externismo (obtenido *a priori*) le permitiría convencerse también de que una persona corpórea y un cerebro en una cubeta no comparten los conceptos relevantes que cada uno de ellos está usando cuando activa el símbolo “cerebro”. Cualquier externista cree (independientemente de creer o no que es un cerebro en una cubeta) que una persona corpórea tiene cierto concepto, digamos CEREBRO-PC, diferente al que tiene un cerebro en una cubeta (si es que éste tiene algún concepto), digamos CEREBRO-CC. Si un externista acepta C, entonces cree que él tiene el concepto CEREBRO-PC y que no puede tener ese otro concepto CEREBRO-CC.²¹ Keanu cree que él tiene el concepto CEREBRO-CC y que no puede tener el concepto CEREBRO-PC. Aunque Keanu acepta la tesis “desentrecomilladora” P₁*, rechaza Q. El aire paradójico de todo el asunto dificulta que podamos hacernos una idea de qué proposición alternativa a Q podría creer Keanu. Consideremos algunas de las opciones que la exposición de Putnam pone sobre la mesa:

Q': El concepto CEREBRO (que uso) tiene en su extensión características concretas de un programa de ordenador.

Q'': El concepto CEREBRO (que uso) tiene en su extensión cerebralmente-aparentes-en-la-“imagen”.

...

²⁰ Ningún rasgo de la situación da a entender que Keanu sea un cerebro en una cubeta. Si lo fuera, todavía sería más complicado describir los presuntos contenidos de sus presuntos estados mentales. Por ello evito usar las posibles connotaciones del nombre “Neox” (de un sujeto que está en *Matrix*), recurriendo al nombre de alguien que en cierto sentido meramente “pretende” ser un cerebro en una cubeta, por el hecho de creerlo (y de alguien que en cierto sentido “pretende” estar en *Matrix*, por el hecho de interpretar a un personaje que está en *Matrix*).

²¹ Usar la expresión “CEREBRO-CC” no implica poder tener el concepto CEREBRO-CC. La expresión “CEREBRO-CC” tiene un carácter descriptivo complejo (cercano al de “el concepto (en caso de existir) que estaría usando un cerebro en una cubeta cuando activara el símbolo interno ‘cerebro’ en situaciones subjetivamente indistinguibles de aquellas en que yo activo el símbolo interno ‘cerebro’”). Análogamente respecto a la expresión “CEREBRO-PC”.

Sea cual sea la hipótesis Q^n creída por Keanu e incompatible con Q , la objeción a la VERSIÓN-A sostiene que la premisa P_2^* sólo resulta plausible si (además de asumir P_2) asumimos que en dicha premisa el término “cerebro” se usa presuponiendo (que su significado se ajusta a) Q ; Keanu rechaza Q , y por lo tanto no podría racionalmente creer que P_2 apoye P_2^* .

Puesto que Keanu acepta P_1 y P_1^* , está obligado a negar P_2^* (dado que rechaza C). Pero —recordemos el ejemplo del chocolate que ilustraba la noción de *petitio*— la cuestión no es si Keanu niega P_2^* , sino si Keanu puede aceptar racionalmente que P_2 proporciona *algún* apoyo a P_2^* . Dicho en otras palabras: ¿resulta más plausible P_2^* (incluso para Keanu, que niega C) en presencia de P_2 de lo que pudiera serlo independientemente de P_2 ?

Creo que la respuesta es positiva. Y aunque en conjunción con Q el apoyo que P_2 confiere a P_2^* es mayor (y más nítido), P_2^* también es apoyada por P_2 en presencia de muchas de esas alternativas que pudiéramos contemplar: Q' , Q'' , etcétera.

Para justificar esa afirmación, tratemos de imaginar la situación epistémica de Keanu en diferentes escenarios: antes y después de llegar a convencerse del externismo encarnado en P_2 .

Antes de aceptar las teorías causal-externistas, las opiniones de Keanu sobre las condiciones de individuación del contenido representacional podrían haber sido muy similares a la que compartirían muchas personas poco familiarizadas con la filosofía y también bastantes filósofos que —por pertenecer a otra época, pongamos por caso— nunca se hubieran planteado las reflexiones que suscitan típicamente nuestras intuiciones externistas (que nunca hubieran oído hablar de Tierras Gemelas, ni siquiera de Wittgenstein) y abrazaran —implícita o explícitamente— lo que usualmente se denomina una concepción *cartesiana* sobre la mente. Keanu y esos otros sujetos podrían ser partidarios de la negación de la tesis (a*):²²

No (a*): *El contenido semántico de una presunta representación (en particular, aquello a lo que refiere) queda determinado por propiedades intrínsecas que comparte con otra presunta representación cualitativamente indistinguible.*

Apliquemos No (a*) a casos concretos. Supongamos que a quien mantiene dicha concepción “imaginística” de la representación se le describe la situación de los cerebros en cubetas. Se le pregunta luego

²² Cfr. la sección 3.

algo como “¿qué estaría pensando un cerebro en una cubeta cuando las características de su vida mental interna fueran subjetivamente indistinguibles de las características de la vida mental interna de un sujeto corpóreo que estuviera pensando que el agua hierve a 100 °C?” La respuesta que considerarían correcta es: “estaría pensando que el agua hierve a 100 °C”. También Keanu, durante esa fase en que aún no es externista, daría una respuesta semejante. De forma similar, unos y otros rechazarían P_2^* , para la que no verían justificación alguna.

Supongamos ahora que esas personas (Keanu entre ellas) llegan a convencerse de P_2 (a través de Wittgenstein, Davidson, Putnam, Burge o por cualquier otra vía). Debe aceptarse que —como señala la objeción de Falvey y Owens conforme a la interpretación que estamos comentando— el impacto de P_2 será diferente, dependiendo de si un sujeto tiene la creencia cotidiana C (cree no ser un cerebro en una cubeta) o, por el contrario, cree No C (el caso de Keanu). El primer grupo de sujetos tiene también la creencia cotidiana Q (adquirir la creencia P_2 no podría ir en detrimento de Q). Así pues, tales sujetos pueden, tras constatar que el símbolo “cerebro” que aparece en el consecuente de P_2^* (también el que aparece en el antecedente) hace referencia a *ciertos objetos físicos determinados del entorno*, emplear P_2 para —siguiendo el camino principal trazado por Putnam— concluir P_2^* (o al menos para considerar que P_2^* está justificada), porque efectivamente creerán que las restricciones histórico-causales requeridas según P_2 dificultan enormemente que un cerebro en una cubeta pudiera *usar (ahora, al tener un pensamiento subjetivamente indistinguible del mío) un concepto cuya extensión está constituida por ciertos objetos físicos determinados del entorno*. Por tanto, se sentirán inclinados a aceptar P_2^* , que considerarán legítimamente sustentada por

Si x fuera un cerebro en una cubeta, entonces x no podría usar (ahora) un concepto cuya extensión está constituida por ciertos objetos físicos determinados del entorno.

Esos sujetos creerán —siguiendo con la línea de Putnam— que, si efectivamente existe ese concepto CEREBRO-CC poseído por los cerebros en cubetas, su extensión estará constituida por características concretas del programa de ordenador, o cerebros-meramente-aparentes-en-la-“imagen”, o algún otro tipo de entidad (que no sea un objeto físico determinado del entorno, en el sentido relevante).

Keanu, por creer No C y rechazar Q, no podrá razonar así. Pero la argumentación de Putnam también ofrece a Keanu (un externista que acepta P_2 pero cree No C) una vía para constatar que resulta difícil sostener la negación de P_2^* (una vía que no estaba abierta cuando Keanu no era externista). Keanu puede creer (como alternativa a Q) que el símbolo “cerebro” que aparece en el antecedente de P_2^* (también el que aparece en el consecuente) hace referencia a características concretas de un programa de ordenador, o a cerebros-meramente-aparentes-en-la-“imagen”, o algo por el estilo. Pero entonces está en disposición de constatar que para que P_2^* sea falso, debería ser verdadero algo como esto:

(#) *Aunque x fuera una característica concreta de un programa de ordenador, o un cerebro-meramente-aparente-en-la-“imagen”, o algo por el estilo, x podría usar (ahora) un concepto cuya extensión es la clase de los cerebros.*

El problema con (#) no deriva de que el símbolo “cerebro” que aparece al final deba interpretarse conforme a Q (Keanu no puede interpretarlo así). El problema es que resultará difícil sostener una teoría que atribuya posesión de conceptos a entidades como las mencionadas en la primera parte de (#), entidades que no son *objetos físicos determinados del entorno*, sino características concretas de un programa de ordenador, o cerebros-meramente-aparentes-en-la-“imagen”, o algo por el estilo. Tengamos en cuenta que para rechazar la objeción de Falvey y Owens (según la interpretación que comentamos) no es preciso suponer que será falsa cualquier teoría que atribuya posesión de conceptos a esas entidades. La acusación de *petitio* en el sentido PJ (Falvey y Owens no especifican otro sentido) establece que *cualquier* sujeto que negara C considerará (si es racional) que hay al menos una premisa que no recibe *ningún* apoyo de los indicios que supuestamente la apoyan. Estamos evaluando la hipótesis de que Keanu (un externista que acepta P_2) debe creer que P_2 no apoya en ningún grado P_2^* . La hipótesis resulta muy dudosa. Cuando Keanu era internista no veía motivo alguno para dudar de la negación de P_2^* (probablemente ni se habría planteado que CEREBRO-PC y CEREBRO-CC tuvieran que ser conceptos diferentes). Ahora, una vez aceptada P_2 , tiene que hacer frente —si quiere seguir negando P_2^* — al reto de acomodar (#) en una teoría plausible sobre la posesión de conceptos. Creo que es razonable describir la situación diciendo que Keanu puede apreciar racionalmente que P_2 confiere cierto apoyo a P_2^* (al constatar que, tras convencerse de P_2 , reconoce en la negación de P_2^* unas dificultades que antes no reconocía).

Ciertamente, la argumentación VERSIÓN-A no tendrá para Keanu el carácter concluyente que Putnam parece pretender que tiene (tampoco queda claro si para los sujetos cotidianos —que creen C— el argumento es suficientemente concluyente). Pero sí tiene cierta fuerza, de tipo abductivo, contra la posición de Keanu. Una fuerza abductiva que no tienen *en absoluto* los ejemplos claros de *petitio*, como MOORE-BIV o CEBRA. Eso es suficiente para negar que haya *petitio* si nos atenemos a la definición PJ.

En los párrafos anteriores indico que no veo razones para afirmar que la VERSIÓN-A prejuzgue la cuestión en el sentido de Jackson. (Eso permite descartar la objeción inspirada en la crítica de Falvey y Owens contra PUTNAM-1, si es que —como he sugerido— es razonable extender dicha crítica a la VERSIÓN-A e interpretar la acusación conforme a PJ.) En las siguientes líneas, hasta el final de esta sección, sostendré que hay margen para defender la VERSIÓN-A incluso si este argumento prejuzgara la cuestión en el sentido de PJ.

Los argumentos clasificados como *petitio* conforme a PJ incurrir en un defecto obvio *solamente si presuponemos* algo que, por defecto, suele ser razonable presuponer (pero que puede no ser de aplicación en algunos casos): *que el destinatario del argumento es quien niega (o quien pone en duda) su conclusión, ya sea de forma real o hipotética*. Si un argumento se dirige a otro destinatario, el hecho de tener la condición descrita en la definición PJ no implica defecto alguno.

No es verdad que la única interpretación razonable del argumento original de Putnam (y de la VERSIÓN-A) sea aceptando esa asunción usual, es decir, aceptando que va dirigido a quien —real o hipotéticamente— niega (o pone en duda) C. Si el argumento se ajustara al esquema PJ (cosa que he rechazado) y ése fuera su destinatario, entonces sería un argumento fallido. Pero cabe imaginar un destinatario diferente. Este destinatario puede ser alguien que no se hubiera planteado inicialmente la cuestión respondida por C. O alguien que habiéndosela planteado, no llega a creer inequívocamente C (sin creer tampoco su negación, ni poner en duda C). O alguien que reconoce creer C, pero tiene dudas sobre el estatus epistémico de dicha creencia: sobre si está suficientemente justificada, por ejemplo. Obtener C como resultado de un argumento con premisas justificadas y de las que se sigue la conclusión es una tarea sensata y potencialmente exitosa si se concibe como respuesta a cualquiera de esos destinatarios.

Sin duda, interpretar de esa forma el argumento (el razonamiento original de Putnam, o nuestra VERSIÓN-A) conlleva imponerle

una importante limitación. Su alcance se vería severamente limitado, pues no aspiraría a poder convencer racionalmente a un opositor. Su importancia filosófica es menor de la que tendría si no existiera dicha restricción. Pero, asumida la limitación, el argumento resulta satisfactorio.

En cualquier caso, no estoy comprometido con esa restricción. Reitero nuevamente la dialéctica seguida en esta sección: la VERSIÓN-A no incurre en *petitio* según la definición de Copi, y tampoco según PJ, la definición de Jackson. He añadido lo siguiente: incluso *si* incurriera en *petitio* según PJ, eso no deslegitima automáticamente al argumento, pues *podría interpretarse* sin asumir que entre sus destinatarios deba incluirse a sujetos que nieguen (o pongan en duda), real o hipotéticamente, la conclusión.

6. La versión basada en la diferencia de contenido

La reinterpretación del argumento que presento y defiendo en esta sección difiere de la VERSIÓN-A por hacer una lectura menos arriesgada de las consecuencias de P_2 . No asume que P_2 garantice la imposibilidad de que el cerebro en la cubeta esté considerando el pensamiento relevante (acerca de cerebros), sino meramente que P_2 garantiza la imposibilidad de que tanto el sujeto corpóreo como el cerebro en la cubeta estén considerando el pensamiento relevante. Eso establecería su premisa P_2' :

VERSIÓN-B:

P_1 : *Estoy considerando el pensamiento de que no soy un cerebro en una cubeta.*

P_2 : *Es verdadera una teoría mínimamente causal y externista sobre el significado.*

P_2' : *Si x fuera un cerebro en una cubeta y z un sujeto corpóreo, entonces x y z no podrían estar considerando cada uno de ellos el pensamiento de que no es un cerebro en una cubeta cuando ambos estuvieran en una situación interna subjetivamente equivalente en la que al menos uno de ellos estuviera considerando el pensamiento de que no es un cerebro en una cubeta.*

P_j : *Si z fuera un sujeto corpóreo, entonces z estaría considerando (ahora, al estar en una situación interna subjetivamente equi-*

valente a la mía) el pensamiento de que no es un cerebro en una cubeta.

P_k: Si x fuera un cerebro en una cubeta y estuviera en una situación interna subjetivamente equivalente a la de un sujeto corpóreo que estuviera (ahora) considerando el pensamiento de que no es un cerebro en una cubeta, entonces x no podría estar considerando (ahora) el pensamiento de que no es un cerebro en una cubeta.

P_n: Si x fuera un cerebro en una cubeta, entonces x no podría estar considerando el pensamiento de que no es un cerebro en una cubeta.

C: (Por lo tanto) No soy un cerebro en una cubeta.

También en esta VERSIÓN-B (como sucedía con la VERSIÓN-A) las premisas P₂ y P_n son innecesarias para derivar C. Las incluyo para visualizar mejor el razonamiento, y su relación con PUTNAM-AC y con la VERSIÓN-A.

La premisa P₂' asevera pues un tipo de tesis mínima, característica del externismo. Una tesis sobre diferencia de contenido-significado. El contenido no viene determinado por los aspectos cualitativos internos compartidos por el sujeto corpóreo y el cerebro en una cubeta. Pero eso no permite concluir C ni siquiera en conjunción con P₁, porque deja abierta la posibilidad de que yo sea un cerebro en una cubeta y (dado que albergo el pensamiento relevante, según establece P₁) a quienes esté vedado tener el pensamiento relevante sea a los sujetos corpóreos. La premisa clave para descartar esa situación es P_j. La premisa P_k viene implicada por P₂', y partiendo de P₁, P_j y P_k podemos concluir C. Así pues, debemos preguntarnos por la justificación de P_j.

Antes, para obtener una visión más nítida de la estructura de esta VERSIÓN-B y facilitar así la constatación de que se mantienen las relaciones que acabo de señalar entre sus diferentes partes, introduzcamos algunas abreviaciones. Estipulemos que “e piensa*” abrevia “e está considerando (ahora) el pensamiento de que e no es un cerebro en una cubeta”. También usaré “f es subjetivamente como g” para abreviar “f y g están en una situación interna subjetivamente equivalente”. Y emplearé la expresión usual “BIV” como abreviatura de “cerebro en una cubeta”. Utilizando esa notación, lo esencial de

la VERSIÓN-B queda formulado de esta otra forma (prescindo ahora de P_2 , que —como sabemos— permite respaldar a P_2'):

VERSIÓN-B:

P_1 : *Yo pienso**.

P_2' : *Si x fuera un BIV y z fuera un sujeto corpóreo, entonces (si x fuera subjetivamente como z y uno de ellos pensara*, entonces no podrían ambos pensar*).*

P_j : *Si z fuera un sujeto corpóreo subjetivamente como yo, entonces z pensaría*.*

P_k : *Si x fuera un BIV y fuera subjetivamente como un sujeto corpóreo que piensa*, entonces x no podría pensar*.*

P_n : *Si x fuera un BIV, entonces x no podría pensar*.*

C : *(Por lo tanto) Yo no soy un BIV.*

Una vez aceptamos que P_j es la premisa crucial en esa argumentación, preguntémonos: ¿cómo se justifica dicha premisa? Ya se ha comentado que tenemos garantía suficiente —por medio del autoconocimiento— sobre la verdad de P_1 . También parece ser cierto que yo soy un sujeto corpóreo (aunque, como veremos en seguida, el objetor de C negaría típicamente esto). Las teorías más plausibles sobre las condiciones de verdad de enunciados subjuntivos como P_j refrendarán (dados algunos otros supuestos empíricos también plausibles) que yo mismo soy un sujeto corpóreo relevantemente similar a mí, del que se cumple el estar teniendo el pensamiento relevante, según establece P_j .

Analícemos cuáles son las dos opciones potencialmente abiertas para un objetor que niegue justificación a P_j , según cómo se posicione en relación con la tesis externista mínima P_2' :

- (i) La motivación escéptica para negar o poner en duda que tuviéramos justificación apropiada para creer C se basaba en la supuesta equivalencia interna subjetiva en que se encontrarían un sujeto corpóreo y un cerebro en una cubeta. La idea era que precisamente por ello no podemos discernir si estamos en una u otra situación. Tradicionalmente, el planteamiento escéptico de ese tipo de escenario estaba vinculado (también en los casos

ideados por Descartes) con la tesis internista de que la diferencia no afecta los contenidos pensados (que se pretenderían equivalentes), es decir, con la falsedad de P_2' .²³ Si P_2' es falso, entonces el sujeto corpóreo y el cerebro en la cubeta (subjétivamente equivalentes) están pensando lo mismo. El autoconocimiento (incorporado en P_1 , aceptable para dicho internista) garantiza que, sea corpóreo o mero cerebro, estoy considerando el pensamiento de que no soy un cerebro en una cubeta. Así, si P_2' es falso, el sujeto corpóreo debe poder considerar dicho pensamiento. Es decir, P_j será verdadera.

(ii) La única vía para el objetor de P_j es aceptar el externismo incorporado en P_2' . Ese sujeto mantendrá, pues, esta combinación de tesis:

– *estoy considerando el pensamiento de que no soy un cerebro en una cubeta* (por P_1);

– *no estoy justificado en creer que un sujeto corpóreo podría considerar dicho pensamiento* (por aceptar P_2' pero objetar a la justificación de P_j);

– *no estoy justificado en creer que soy un sujeto corpóreo* (pues teniendo dicha justificación y la justificación del autoconocimiento relevante, podría justificar P_j).

El camino por el que hemos llegado a suscitar esa posición puede hacernos pensar que dicha combinación es algo *ad hoc* y poco natural. Probablemente se replicará que la posición (perfectamente inteligible y consistente) puede generarse y justificarse por otras vías. Creo, con todo, que la mejor manera de atacar esa posición es basándonos en consideraciones de simplicidad. La combinación de esas tres tesis resulta menos simple, menos natural, que la combinación de P_1 , P_2' y la tesis de que estoy justificado en creer que soy un sujeto corpóreo.

²³ Tal y como ha resaltado un árbitro anónimo de *Crítica* en sus comentarios a una formulación previa de esta discusión, el vínculo entre el planteamiento de la duda escéptica (sobre si tenemos justificación para C dado que supuestamente no podemos discernir subjétivamente si somos corpóreos o no) y las teorías internistas del contenido es contingente, de forma que es compatible una posición externista con el planteamiento de esa duda escéptica. Sus críticas me han hecho ver los puntos débiles de la VERSIÓN-B como reconstrucción del argumento original de Putnam, aunque considero que (complementado con las consideraciones de simplicidad que menciono a continuación en el texto) es un buen argumento.

Posiblemente ese juicio comparativo sobre simplicidad dependa de considerar que el planteamiento de la duda escéptica (basada en la indiscernibilidad subjetiva de los dos escenarios) es más afín a la concepción internista que a la concepción externista sobre el contenido (aunque también sea lógicamente compatible con esta última, como se ha reconocido en el párrafo anterior). Si, por el contrario, toda la verdad de ese juicio comparativo sobre simplicidad dependiera meramente de que la hipótesis de que soy corpóreo es más simple que su negación, entonces la VERSIÓN-B sería dudosa como reconstrucción y defensa del argumento original de Putnam, porque haría depender excesivamente dicho argumento de esa cuestión (la hipótesis de que soy corpóreo es más simple que su negación), implicando con ello que los otros elementos son comparativamente casi superfluos. Pero no creo que nuestra tendencia a rechazar la posición escéptica contraria a P_j dependa sólo de eso.

Esta VERSIÓN-B requiere, en cualquier caso, invocar esas consideraciones de simplicidad que aconsejan abandonar la posición escéptica que se acaba de describir. El hecho de que no aparezcan tan explícitamente mencionadas por Putnam,²⁴ también apoya que la VERSIÓN-A es exegéticamente preferible.

7. Conclusiones

Resumiré en esta sección los puntos principales desarrollados. El argumento de Putnam está destinado a mostrar que es epistémicamente imposible que seamos cerebros en una cubeta (es decir, sabemos que no lo somos). La formulación PUTNAM-1 recoge lo esencial de dicho argumento, resaltando (en su premisa P_1) el papel desempeñado por el autoconocimiento de los propios pensamientos. La versión PUTNAM-AC incorpora explícitamente una premisa externista (P_2) que resume dos proposiciones empíricas de las que —según admite Putnam— depende su argumentación. Partiendo de esa versión, todavía considerablemente abstracta, he especificado y defendido dos reinterpretaciones distintas del argumento, que difieren en su interpretación de las consecuencias de P_2 . La VERSIÓN-A extrae de P_2 la consecuencia de que el cerebro en una cubeta no puede tener (cuando está en una situación subjetivamente indistinguible respecto de la de un sujeto corpóreo) el pensamiento relevante (acerca de cerebros) que P_1 garantiza que tenemos. La VERSIÓN-B extrae de P_2 sólo la consecuencia de que un sujeto corpóreo y un cerebro en

²⁴ No obstante, *cfr.* sus comentarios en Putnam 1981, p. 16.

una cubeta no pueden ambos tener (en situación de equivalencia subjetiva) el pensamiento relevante. Como eso no bastaría para alcanzar la conclusión, incorpora también una premisa respaldada por proposiciones cotidianas, estableciendo que un sujeto corpóreo puede tener el pensamiento relevante.

He rechazado una posible objeción a la VERSIÓN-A inspirada por una crítica formulada por Falvey y Owens contra PUTNAM-1. (Se obtendrían conclusiones análogas si la crítica se dirigiera contra la VERSIÓN-B.) El objetor acusa al argumento de prejuzgar la cuestión por depender de proposiciones empíricas que sólo podrían justificarse aceptando previamente su conclusión. Asumiendo la definición tradicional de Copi, la VERSIÓN-A no incurre en *petitio*. Obtenemos el mismo resultado negativo (aunque no resulta tan obvio) si asumimos una caracterización posterior de la *petitio*, propuesta por Jackson. Además, aunque el argumento prejuzgara la cuestión en el sentido de Jackson, eso sólo lo deslegitimaría si suponemos que su destinatario es quien niega (o pone en duda), real o hipotéticamente, la conclusión; pero es factible interpretarlo prescindiendo de esa suposición.

El punto crucial del que depende distintivamente la VERSIÓN-B es que un sujeto corpóreo puede tener el pensamiento relevante (sobre cerebros), tesis implicada por proposiciones de sentido común. Poner en duda esa premisa entra en conflicto con la posición internista generalmente vinculada con la motivación natural inicial que pudiera haber existido para dudar de nuestra justificación para creer la conclusión del argumento de Putnam; mantener dicha motivación en un marco externista es lógicamente consistente pero menos simple que, desde el externismo, aceptar que un sujeto corpóreo puede tener el pensamiento relevante (sobre cerebros).*

*He expuesto una parte de este trabajo en varios foros científicos. Algunas ideas de las secciones 2, 3 y 8 (recogidas en la publicación Pérez Otero 2010a) se presentaron en el VI Congreso de la SEFA (*Sociedad Española de Filosofía Analítica*), celebrado en Santa Cruz de Tenerife, en octubre de 2010. Esos contenidos y otras consideraciones que ampliaban el marco de la discusión se exponían en el *Taller de Lógica y Argumentación* del XI Coloquio Compostelano de Filosofía Analítica, organizado en la Universidad de Santiago de Compostela, en junio de 2011. Agradezco a los diversos participantes en esas reuniones sus comentarios y sugerencias críticas, en especial a María José Frápolli, Tobies Grimaltos y Carlos Moya. También estoy en deuda con un árbitro anónimo de *Crítica*, cuyos comentarios han contribuido a clarificar y corregir parcialmente algunos puntos. Financiación: Programa CONSOLIDER-INGENIO 2010, “Perspectival Thoughts and Facts” (CSD2009-00056), MICINN (Gobierno de España). Proyecto “Conocimiento, referencia y realismo” (FFI2011-29560-C02-01), MICINN. Grupo de investigación consolidado *LOGOS* (2009 SGR 1077), AGAUR (Generalitat de Catalunya).

BIBLIOGRAFÍA

- Burge, Tyler, 1979, "Individualism and the Mental", *Midwest Studies in Philosophy*, vol. 4, pp. 73–121.
- Copi, Irving M., 1961 (1978), *Introduction to Logic*, 5a ed., Macmillan, Nueva York (a cuya paginación remito). [Versión en castellano: *Introducción a la lógica*, 8a. ed., trad. Edgar A. González, Limusa, México, 1995.]
- Davies, Martin, 2009, "Two Purposes of Arguing and Two Epistemic Projects", en Ian Ravenscroft (comp.), *Minds, Worlds and Conditionals: Essays in Honour of Frank Jackson*, Oxford University Press, Oxford, pp. 337–383. También disponible en: <http://www.mkdavies.net/Martin_Davies/Epistemology_files/TwoProjects.pdf> [última consulta: 14/11/2012].
- Dretske, Fred, 1970, "Epistemic Operators", *Journal of Philosophy*, vol. 67, pp. 1007–1023.
- Falvey, Kevin y Joseph Owens, 1994, "Externalism, Self-Knowledge, and Skepticism", *Philosophical Review*, vol. 103, pp. 107–137.
- Hazlett, Allan, 2006, "Epistemic Conceptions of Begging the Question", *Erkenntnis*, vol. 65, pp. 343–363.
- Jackson, Frank, 1987, "Petitio and the Purpose of Arguing", *Conditionals*, Basil Blackwell, Oxford, pp. 100–114. (Versión revisada del artículo homónimo publicado en *Pacific Philosophical Quarterly*, vol. 65, 1984, pp. 26–36.)
- Moore, George E., 1939, "Proof of an External World", *Proceedings of the British Academy*, vol. 25. (Reimpreso en G.E. Moore, *Philosophical Papers*, George Allen and Unwin, Londres, 1959, pp. 126–148.)
- Pérez Otero, Manuel, 2012, "Los propósitos de razonar, ilustrados con el argumento externista anti-escéptico de Putnam", *Theoria*, vol. 27, pp. 55–74.
- , 2011, "Modest Skepticism and Question Begging Proper", *Grazer Philosophische Studien*, vol. 83, pp. 9–32.
- , 2010a, "Defensa de una versión del argumento trascendental de Putnam contra la posibilidad epistémica de que seamos cerebros en cubetas", en A. Jaume, M. Liz, D. Pérez Chico, M. Ponte y M. Vázquez (comps.), *Proceedings of the Sixth Conference of the Spanish Society for Analytic Philosophy*, Santa Cruz de Tenerife, pp. 139–140.
- , 2010b, "El debate entre externistas e internistas sobre la racionalidad inferencial", *Análisis Filosófico*, vol. 30, pp. 163–186.
- , 2009a, "Transmisión de la justificación epistémica", en Daniel Quesada (coord.), *Cuestiones de teoría del conocimiento*, Tecnos, Madrid, pp. 148–180.
- , 2009b, "Conocimiento, discriminabilidad y acceso al contenido representacional", en J. Alcolea, V. Iranzo, A. Sánchez y J. Valor (comps.), *Actas del VI Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*, Universitat de València, Valencia, pp. 227–231.

- Pérez Otero, Manuel, 2006, "Limitaciones del dogmatismo como réplica ante el reto escéptico modesto", en F. Martínez Manrique y L.M. Perís-Viñé (comps.), *Actas del V Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*, editorial, Granada, pp. 185–189.
- , 2004a, "Las consecuencias existenciales del externismo", *Análisis Filosófico*, vol. 24, pp. 29–58.
- , 2004b, "El externismo intencional ante la transparencia de las actitudes proposicionales", en A. Vicente, P. de la Fuente, C. Corredor, J. Barba y A. Marcos (comps.), *Actas del IV Congreso de la Sociedad de Lógica, Metodología y Filosofía de la Ciencia en España*, editorial, Valladolid, pp. 262–265.
- (inédito a), "Purposes of Reasoning and (a New Vindication of) Moore's Proof of an External World".
- (inédito b), "Cogent Argumentation without Transmission of Warrant. (The Case of Moore's Proof)".
- Pryor, James, 2004, "What's Wrong with Moore's Argument?", *Philosophical Issues*, vol. 14, pp. 349–378.
- Putnam, Hilary, 1981, "Brains in a Vat", *Reason, Truth and History*, Cambridge University Press, Cambridge, pp. 1–21.
- , 1975, "The Meaning of 'Meaning'", *Mind, Language and Reality*, Cambridge University Press, Cambridge, 1975, pp. 215–271.
- , 1973, "Meaning and Reference", *The Journal of Philosophy*, vol. 70, pp. 699–711.
- Quesada, Daniel, 1998, *Saber, opinión y ciencia: Una introducción a la teoría del conocimiento clásica y contemporánea*, Ariel, Barcelona.
- Wright, Crispin, 2000, "Cogency and Question-Begging: some Reflections on McKinsey's Paradox and Putnam's Proof", *Philosophical Issues*, vol. 10, pp. 140–163.
- , 1994, "On Putnam's Proof That We Are Not Brains in a Vat", en Peter Clark y Bob Hale (comps.), *Reading Putnam*, Blackwell, Oxford, pp. 216–241.

Recibido el 19 de julio de 2011; revisado el 12 de marzo de 2012; aceptado el 23 de marzo de 2012.