

El autoconocimiento de las creencias: una objeción al método de la transparencia

Beliefs' self-knowledge: an objection to the method of transparency

Javier Vidal

Universidad de Concepción, Chile
fravidal@udec.cl

Resumen

Según el método de la transparencia, el autoconocimiento genuino es el producto de una inferencia del mundo a la mente. A. Byrne (2018) ha desarrollado una teoría en la que el método de la transparencia consiste en seguir una regla epistémica a fin de formarse creencias de segundo orden que sean autoverificantes. En este artículo, argumento que la teoría de Byrne no establece condiciones suficientes para tener autoconocimiento de las creencias de primer orden. Examinando un caso de autoengaño, trato de mostrar que seguir tal regla puede no resultar en autoconocimiento cuando uno está involucrado en la deliberación racional. En el caso bajo consideración, precisamente uno llega a creer que uno cree que p sin llegar a creer que p . La justificación de que uno no se forme la creencia de que p con su patrón causal característico en la vida mental y el comportamiento es que uno ya tenía la creencia inconsciente de que no p , la cual no es sensible a los principios que gobiernan el razonamiento teórico y práctico.

Palabras clave: autoengaño, creencia inconsciente, creencia de segundo orden, deliberación racional, regla doxástica.

Abstract

According to the method of transparency, genuine self-knowledge is the outcome of an inference from world to mind. A. Byrne (2018) has developed a theory in which the method of transparency consists in following an epistemic rule in order to form self-verifying second-order beliefs. In this paper, I argue that Byrne's theory does not establish sufficient conditions for having self-knowledge of first-order beliefs. Examining a case of self-deception, I strive to show that following such a rule might not result in self-knowledge when one



Received: 18/09/2019. Final version: 12/12/2019

eISSN 0719-4242 – © 2019 Instituto de Filosofía, Universidad de Valparaíso

This article is distributed under the terms of the

Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Internacional License



CC BY-NC-ND

is involved in rational deliberation. In the case under consideration, one precisely comes to believe that one believes that p without coming to believe that p . The justification for one's not forming the belief that p with its distinctive causal pattern in mental life and behaviour, is that one already had the unconscious belief that not- p , a belief that is not sensitive to the principles governing theoretical and practical reasoning.

Keywords: self-deception, unconscious belief, second-order belief, rational deliberation, doxastic rule.

1. Introducción

Podemos llegar a saber que tenemos una creencia o un deseo mediante una inferencia a partir de la observación de la conducta o quizá, con la ayuda de la psicoterapia, a través de la interpretación de ciertos síntomas. Puedo llegar a saber, por ejemplo, que yo creo que tirar la sal trae mala suerte como resultado de observar mi conducta cuidadosa en las comidas. En estos casos el conocimiento de nuestros estados mentales no es distinto del conocimiento que de ellos pueden tener los demás, y no parece que los métodos para el autoconocimiento aquí involucrados proporcionen alguna garantía especial frente al error. Diríamos entonces que estos no son casos de autoconocimiento en el sentido genuino que discutiremos en el presente artículo. En efecto, seguiré a A. Byrne en plantear el problema del autoconocimiento, respecto a ciertos estados mentales, como el problema consistente en conciliar el carácter *peculiar* y el carácter *privilegiado* que es propio del conocimiento que ordinariamente tenemos de tales estados, o del acceso a ellos (Byrne 2018, 5-9; Fricke 2019, 181-182). Cuando tengo la creencia de que yo creo que está lloviendo, cuando tengo esa creencia de segundo orden, el autoconocimiento es privilegiado porque el error es muy poco probable (a diferencia de lo que quizá ocurre con la creencia de que está lloviendo). Además, en este caso el autoconocimiento es peculiar en el sentido de que parece poseer un carácter directo e inmediato que no comparte el conocimiento que los demás tienen de que yo creo que está lloviendo. Ahora bien, ¿por qué hablar de un *problema* del autoconocimiento generado por la presencia conjunta de la peculiaridad y el privilegio? La razón es que la peculiaridad y el privilegio son aspectos independientes entre sí, puesto que podría haber un autoconocimiento privilegiado que no fuera peculiar y también podría haber un autoconocimiento peculiar que no fuera privilegiado (Byrne 2018, 9-10)¹. Más aún, parece haber cierta tensión entre ambos aspectos, lo que muestra el hecho, externo a la esfera del autoconocimiento, de que el conocimiento científico privilegiado, como en astronomía, suele ser precisamente un conocimiento que debe su

¹ El punto de Byrne es señalar que, en la medida en que ciertas teorías del autoconocimiento solo podrían explicar uno de los dos aspectos (la teoría inferencialista de Ryle y las teorías del sentido interno, respectivamente), no son teorías satisfactorias.

privilegio al uso de instrumentos y métodos de observación cuyos resultados pueden ser contrastados por los distintos investigadores (es decir, es un conocimiento que debe su privilegio a la falta de peculiaridad).

En este artículo, me propongo articular una objeción al método de la transparencia, que, basándose en algunas consideraciones por parte de G. Evans (1982), es una de las estrategias más recientes y extendidas para solucionar el problema del autoconocimiento (Gallois 1996; Moran 2001; Fernández 2013; Byrne 2018). Me ocuparé específicamente de la teoría de Byrne. Aunque este autor ha desarrollado ambiciosamente una teoría de la transparencia para explicar el autoconocimiento genuino de creencias, deseos, intenciones y estados perceptuales, solo examinaré el caso de las creencias, que parece ser el caso más sólido y el que ha recibido más atención. Empezaré describiendo brevemente el método de la transparencia en la versión de Byrne (sección 2). A continuación, presentaré el caso de autoengaño que guiará la discusión, primero en términos de una crítica a cierta interpretación de lo que está ocurriendo (sección 3). Considerando la naturaleza de las creencias de primer y segundo orden, respectivamente, argumentaré después por qué ese es un caso de autoengaño en el que, a través de la deliberación racional, el sujeto llega a formarse una creencia de segundo orden sin formarse también la correspondiente creencia de primer orden (sección 4). Estaré así en condiciones de mostrar que la teoría de Byrne no puede establecer condiciones suficientes para tener un autoconocimiento genuino de las creencias (sección 5).

2. El método de la transparencia: la teoría de Byrne

Cuando se habla de transparencia en la literatura, es frecuente referirse a la idea de que una clase de estados mentales es tal que un sujeto no puede encontrarse en un estado de esa clase sin tener autoconocimiento genuino. Podría argumentarse, por ejemplo, que es una necesidad o verdad conceptual que si, en condiciones normales, un sujeto tiene una creencia de que p , entonces tiene la creencia de que él cree que p (Wright 2001; Bilgrami 2006; Coliva 2016). Sin embargo, no es esta la noción de transparencia que es relevante aquí. El método de la transparencia aplicado a las creencias consiste en llegar a conocer qué creencia tiene uno por dirigir la atención a los mismos hechos del mundo que le llevan a tener esa creencia en primer lugar. Se trata de conocer las propias creencias “mirando hacia fuera”, por así decirlo, en vez de realizando un ejercicio introspectivo. Para responder a la pregunta “¿Crees que habrá una tercera guerra mundial?” no examino ningún hecho interno: tengo que prestar atención a los mismos hechos externos a los que atendería si tuviera que contestar a la pregunta “¿Habrá una tercera guerra mundial?”. En otras palabras, estoy en posición de contestar a la pregunta de si creo que p poniendo en marcha el procedimiento (cualquiera que este sea) mediante el cual respondo a la pregunta de si p (Evans 1982, 225).

En la medida en que se trata de una *inferencia* del mundo a la mente, parezco estar siguiendo el esquema doxástico (Gallois 1996, 47):

$$\frac{p}{\text{Creo que } p}$$

Byrne desarrolla esta idea en términos de la noción de regla epistémica (Byrne 2018, 100-102). Por ejemplo, un sujeto llega a formarse creencias (y así adquirir conocimiento) de las visitas a su casa siguiendo esta regla doxástica:

(TIMBRE) Si suena el timbre, cree que hay alguien en la puerta.

El sujeto reconoce que suena el timbre mediante la percepción, es decir, escuchando el timbre, y luego, siguiendo esa regla, llega a creer (y quizá saber) que hay alguien en la puerta mediante la correspondiente inferencia. Pues bien, la regla doxástica para la formación de creencias de segundo orden, y así adquirir autoconocimiento de las creencias (de primer orden), es:

(CRE) Si p , cree que tú crees que p .

Seguir esta regla en un caso normal consistiría, en primer lugar, en reconocer que p es el caso, lo que implica juzgar o creer que p (cuando de hecho p es el caso). A continuación, el sujeto se forma inferencialmente la creencia de que él cree que p . A la formulación lingüística de la regla en la forma de un condicional le corresponde psicológicamente el hecho de que el sujeto llega a creer que él cree que p *porque* reconoce que p . Este “porque” indica el seguimiento de la regla en términos de razones con base en una conexión causal entre la creencia de primer orden y la creencia de segundo orden. Ciertamente, un sujeto puede creer que p cuando p no es el caso. En esa circunstancia no puede decirse que el sujeto siga la regla, puesto que seguir la regla requiere el reconocimiento (o conocimiento) de que p es el caso. La idea entonces es que el sujeto *trata* de seguir la regla sin conseguirlo (Byrne 2018, 111), lo que no tiene consecuencias relevantes para la formación de la creencia de segundo orden: igualmente ocurre que creer que p es la causa de llegar a creer que él cree que p .

Pordemos ver ahora en qué sentido la teoría de Byrne explicaría el autoconocimiento peculiar y privilegiado de nuestras creencias (Byrne 2018, 108-109). Por una parte, es claro que un sujeto no puede usar una regla doxástica análoga para llegar a conocer las creencias de *otro* sujeto, digamos S. En efecto, consideremos la regla:

(CRE*) Si p , cree que S cree que p .

Que yo reconozca (o crea falsamente) que p es el caso no constituye ninguna razón para creer que un sujeto S distinto de mí cree que p . Es absurdo pensar que mis creencias

sobre el mundo me permitan conocer qué creencias tienen los demás. Por lo tanto, para conocer qué creencias tienen los demás debemos usar algún método diferente al método de la transparencia. Así, en la medida en que, para conocer mis creencias, solo yo puedo usar una regla doxástica que me lleva a realizar una inferencia del mundo a la mente, el autoconocimiento (o acceso) resultante es genuinamente peculiar. Ciertamente, seguir la regla (CRE) es usar un procedimiento inferencial, pero tratándose de una inferencia simple esto no excluye el carácter directo e inmediato del autoconocimiento en comparación con el conocimiento que tenemos de las creencias de otros a través de la observación de su conducta y de inferencias más complejas (Fricke 2019, 213). Por otra parte, no cabe duda de que el autoconocimiento resultante es privilegiado, puesto que, en general, las creencias de segundo orden que llego a formarme se *autoverifican*. Para seguir (o tratar de seguir) la regla (CRE) debo tener la creencia de que p , pero justamente tener esa creencia es lo que hace verdadera la creencia de segundo orden que llego a formarme inferencialmente. Veámoslo de este otro modo. Cuando siguiendo la regla doxástica me formo la creencia de segundo orden, hay, en general, una transición causal entre mi creencia de que p y mi creencia de que yo creo que p que garantiza la verdad de esa creencia. Seguir la regla (CRE*) no garantiza la verdad de mi creencia de que S cree que p porque, como es obvio, esa creencia de segundo orden tiene una conexión causal con *mi* creencia de que p , no con la creencia de S (como debería ser para que la transición causal garantizase la verdad).

Me referiré brevemente ahora a una crítica a Byrne que es relevante para el desarrollo del artículo. Se ha argumentado que seguir la regla doxástica solo puede explicar el autoconocimiento genuino de las creencias que uno llega a adquirir precisamente por poner en marcha este método, no así el autoconocimiento genuino de las creencias que uno *ya* tenía previamente (Bar-On 2004, 118-119; Gertler 2011). Puesto que la idea original es que respondo a la pregunta “¿Crees que habrá una tercera guerra mundial?” como resultado de responder a la pregunta “¿Habrá una tercera guerra mundial?”, parece que aplicar el método de la transparencia supone evaluar (o re-evaluar) la evidencia disponible y, por lo tanto, involucrarse en un proceso de deliberación racional acerca de si habrá o no una tercera guerra mundial. Llegar a una conclusión afirmativa me llevaría *ahora* a adquirir la creencia correspondiente y luego a formarme la creencia de que yo creo que habrá una tercera guerra mundial. Pero supongamos que antes de la deliberación yo tenía la creencia consciente de que no habrá una tercera guerra mundial. El problema es que no parece que seguir la regla doxástica sea el método que pueda explicar el autoconocimiento genuino de esa creencia. La respuesta de Byrne consiste en desligarse de la formulación interrogativa de Evans, distinguiendo entre el método de la transparencia y la deliberación racional. Si ya tengo la creencia consciente de que no habrá una tercera guerra mundial, me formo inferencialmente la correspondiente creencia de segundo orden sin evaluar ninguna evidencia. Trato de seguir la regla doxástica sin haberme involucrado en ninguna deliberación racional. Más tarde puedo decidir darle otra vuelta a la cuestión

e involucrarme en una deliberación que termine en el juicio (y así en la creencia) de que habrá una tercera guerra mundial, lo que implica un cambio de creencias. Tratando de seguir la regla doxástica me formo ahora la creencia de que yo creo que habrá una tercera guerra mundial (Byrne 2018, 125-126)².

3. Un caso de autoengaño en disputa

Me propongo presentar una objeción a la teoría de la transparencia de Byrne, para lo cual consideraré un caso especial de autoengaño. Se trata del caso de un profesor o profesora que, aunque juzga con fundadas razones que los estudiantes extranjeros son tan buenos como los nacionales, resulta evidente por decisiones académicas que toma, como hacer recomendaciones, que en realidad no tiene esa creencia en absoluto (Peacocke 1999, 242)³. De entrada, parece claro, o al menos daré por hecho, que este es un caso en el que el sujeto tiene la creencia de segundo orden de que él cree que los estudiantes extranjeros son tan buenos como los nacionales (pero volveré sobre esto más adelante). Pues, por ejemplo, estará dispuesto a aseverar sinceramente la oración “Creo que los estudiantes extranjeros son tan buenos como los nacionales”. En el propio planteamiento de Peacocke realizar un juicio consciente de que p es tener una razón para autoadscribirse la creencia de que p , de tal modo que en condiciones normales esa autoadscripción, o creencia de segundo orden, es verdadera (y, más aún, produce autoconocimiento). Esto es así porque existe una relación constitutiva entre creencia y juicio tal que normalmente realizar un juicio de que p , que es un acto mental ocurrente, es un modo de formarse la creencia correspondiente para retener el contenido juzgado previamente y así poder utilizarlo posteriormente. En este caso simplemente ocurriría que no se dan las condiciones normales para formarse la creencia de que p y que, por lo tanto, la creencia de segundo orden es falsa (Peacocke 1999, 237-243). Sin embargo, Peacocke no da una explicación de lo que hace que en este caso las condiciones no sean normales.

Una primera aproximación a esta explicación es proponer que, a diferencia de lo que ocurre normalmente, el profesor ya tiene la creencia inconsciente de que los estudiantes extranjeros *no* son tan buenos como los nacionales (es decir, ya tiene la creencia de que no p). De hecho, no parece que tomar decisiones académicas sesgadas, en el sentido de favorecer a los estudiantes nacionales cuando la evidencia disponible no los destaca, pueda deberse solamente al no tener la creencia de que p , de lo que más bien se esperaría una conducta más neutral. Ahora bien, podría argumentarse que, aunque ciertamente puede

² Esta crítica podría completarse con la objeción al método de la transparencia según la cual este método no puede explicar el autoconocimiento genuino de creencias *recalcitrantes*, esto es, creencias que uno mantiene a pesar de llegar a saber que no hay ninguna razón o evidencia para creer (Cassam 2014, 106-111). Conjeturo que Byrne respondería del mismo modo.

³ Es importante tener en cuenta que, *ex hypothesi*, este es un caso en el que el sujeto tiene excelentes razones o evidencia para juzgar que los estudiantes extranjeros son tan buenos como los nacionales.

explicarse así el comportamiento discriminatorio del sujeto, esto es distinto de explicar por qué el juicio de que p no lleva a la formación de la creencia de que p . (Mi posición es que no puede hacerse esta distinción entre ambas explicaciones, como se verá más adelante).

Más aún, es posible, una vez que se introduce esta distinción, argumentar frente a Peacocke que el sujeto sí llega a formarse la creencia de que p , a pesar de seguir teniendo la creencia de que no p . Resultaría entonces que, después de juzgar que p , el sujeto tiene creencias inconsistentes. Este planteamiento no está motivado por la relación constitutiva entre juicio y creencia sino por la tesis *constitutivista* según la cual las creencias de segundo orden configuran nuestra mente. Esto significa, en algunas versiones del constitutivismo (Wright 2001; Bilgrami 2006; Coliva 2016), que es una verdad o necesidad conceptual que si, en ciertas condiciones normales, un sujeto cree que él cree que p , entonces cree que p ⁴. Justamente vimos que en el caso descrito por Peacocke es poco discutible que el profesor tiene la creencia de que él cree que los estudiantes extranjeros son tan buenos como los nacionales. Se seguiría, por tanto, que él tiene también la creencia de primer orden. Pues, en principio, no habría en este caso un problema con las condiciones normales consideradas por un constitutivista como Wright, que hacen referencia básicamente a la lucidez cognitiva y la posesión de los conceptos relevantes (el concepto de creencia y el concepto de primera persona) en la formación de la creencia de segundo orden. (Sin embargo, mi crítica posterior versará sobre este punto)⁵. ¿Cómo puede entonces el constitutivista sostener convincentemente que en este caso el sujeto tiene creencias inconsistentes, sin cuestionar su racionalidad? La propuesta desarrollada por Coliva y Bilgrami consiste en proponer una distinción entre creencias de primer orden (y entre estados intencionales en general): por un lado, estarían las creencias como *compromisos*, o sensibles al juicio, que son las creencias producidas por la deliberación y de las que el sujeto es racionalmente responsable; y, por otro lado, estarían las creencias como *disposiciones*, que son las creencias sustraídas al control directo del sujeto, por lo que no es racionalmente responsable de ellas (Coliva 2016, 26-34; Bilgrami 2006). Se trata así, en el caso del profesor, que la creencia de que no p es un estado disposicional que tiene una dinámica cognitiva independiente en relación al sujeto, lo que se traduce en ciertos comportamientos discriminatorios. En cambio, su deliberación acerca de qué creer resulta en un compromiso, la creencia de que p , mediante el cual el sujeto se somete a los principios que gobiernan el razonamiento teórico y práctico (Coliva 2016, 198-200; Bilgrami 2006).

⁴ Mientras que este condicional representa la *autoridad* (o infalibilidad) de la primera persona, el condicional inverso representa la *transparencia* en el sentido que no es relevante en el artículo, como señalé al comienzo de la sección 2. Shoemaker propone una versión distinta del constitutivismo en la que el profesor también llegaría a creer que p (Shoemaker 2009, 43-44).

⁵ Es cierto que Wright incluye entre las condiciones normales el hecho de que el sujeto no haya caído en el autoengaño, pero no está claro qué casos de autoengaño está considerando (Wright 2001, 324). Volveré sobre este punto más adelante.

Este planteamiento podría ser aceptable si la distinción entre creencias como compromisos y como disposiciones correspondiese a una distinción entre creencias ocurrentes y creencias permanentes. Recuérdese que en la práctica el profesor se comportará en términos de la creencia de que no p incluso después de la deliberación y el juicio resultantes en la creencia de que p . Si la creencia de que p fuese una creencia ocurrente, en el sentido preciso de no sobrevivir al juicio para retener el contenido juzgado (y poder utilizarlo posteriormente), eso explicaría que el sujeto se siguiera comportando en términos de la creencia de que no p . Pero el concepto de creencia, como diferente del concepto de juicio, parece excluir esta posibilidad: una creencia no es un acto (o acción) mental cuyos efectos puedan variar en el tiempo (Cassam 2014, 117-118). Coliva es explícita en rechazar esta posibilidad (Coliva 2016, 28), lo cual es claro, por otra parte, dado que en su propuesta las creencias como compromisos y como disposiciones son igualmente creencias permanentes, en el sentido de ser estados que tienen efectos duraderos tanto en la vida mental como en el comportamiento del sujeto. La diferencia radica solamente en si el contenido retenido está o no disponible para que el sujeto pueda guiar su razonamiento y su acción. ¿Cuál sería entonces el razonamiento o acción que manifiesta la creencia de que p , cuando el sujeto sigue comportándose básicamente en términos de la creencia de que no p ? Podría sugerirse en primera instancia que se trata de la disposición a juzgar ese contenido, o a aseverar sinceramente la oración “Los estudiantes extranjeros son tan buenos como los nacionales”. Ahora bien, de tener esa disposición no se sigue que el sujeto tenga la creencia correspondiente. Como ya hemos visto, no está en cuestión por el momento que el profesor tiene la creencia de segundo orden que le dispone a aseverar sinceramente la oración “Creo que los estudiantes extranjeros son tan buenos como los nacionales”. Pero puede argumentarse que, a pesar de poseer distintas condiciones de verdad, las oraciones de la forma “ p ” y “Creo que p ” poseen las mismas condiciones de uso, o son performativamente equivalentes, en el sentido de que si uno realiza una aserción de la forma “ p ” está dispuesto a realizar una aserción de la forma “Creo que p ”, y viceversa. Esto se justifica porque, en la medida en que una oración de la forma “ p ” expresa normalmente una creencia (o pensamiento) consciente, el sujeto debe tener la correspondiente creencia de segundo orden, lo que le coloca en posición de reportar esa creencia (o pensamiento) consciente mediante una aserción de la forma “Creo que p ” (Rosenthal 2005, 269-272, 294-305). Ahora bien, una vez que el sujeto ha adquirido ese hábito, nada impide que, sin tener la creencia consciente de que p , la creencia de que él cree que p sea suficiente para colocarle en posición de aseverar tanto “Creo que p ” como “ p ”. Si esto es correcto, es suficiente que el profesor tenga la creencia de segundo orden (sin tener la creencia de primer orden) para estar dispuesto a aseverar la oración “Los estudiantes extranjeros son tan buenos como los nacionales”⁶.

⁶ Esto no significa que no haya excepciones a esa equivalencia performativa. Cuando en un caso freudiano de creencia inconsciente, un sujeto llega a formarse una creencia de segundo orden a través de la interpretación psiconalítica, estará dispuesto a aseverar la correspondiente oración de la forma “Creo que p ”. Pero

Mi argumentación ahora estará destinada, en primer lugar, a mostrar que tener la creencia inconsciente de que no p es lo que explica tanto el comportamiento discriminatorio del sujeto como, precisamente por ello mismo, el hecho de no formarse la creencia de que p después de una deliberación que termina en el juicio correspondiente. Así, estaré de acuerdo con Peacocke en que el juicio de que p no lleva a formarse la creencia de que p en este caso pero, a diferencia de él, propondré una explicación de lo que está ocurriendo. Estaré de acuerdo con Coliva y Bilgrami en que la creencia de que no p es lo que explica el comportamiento discriminatorio del sujeto pero, a diferencia de ellos, daré una explicación de este caso de autoengaño que no requiere atribuir al sujeto tanto la creencia de que p como la creencia de que no p . Esta explicación, que es más económica y, en la medida en que no requiere atribuir creencias inconsistentes al sujeto, introduce un menor grado de irracionalidad, está además en la línea de las teorías deflacionarias o no intencionalistas de los casos cotidianos de autoengaño (Mele 2000; Fernández Acevedo 2018). Trataré, en segundo lugar, de argumentar por qué en este caso el sujeto se forma una creencia de segundo orden, a pesar de no llegar a formarse la correspondiente creencia de primer orden. No me contentaré, por lo tanto, con dejar constancia de que el sujeto se forma la creencia de segundo orden, como he supuesto hasta aquí, sino que daré una explicación de ese hecho en términos de la relación entre creencias y conducta. Estaré así en condiciones de mostrar, finalmente, que el método de la transparencia, en particular la versión de Byrne consistente en seguir la regla doxástica (CRE), no establece condiciones suficientes para el autoconocimiento genuino.

4. Los límites de la deliberación racional

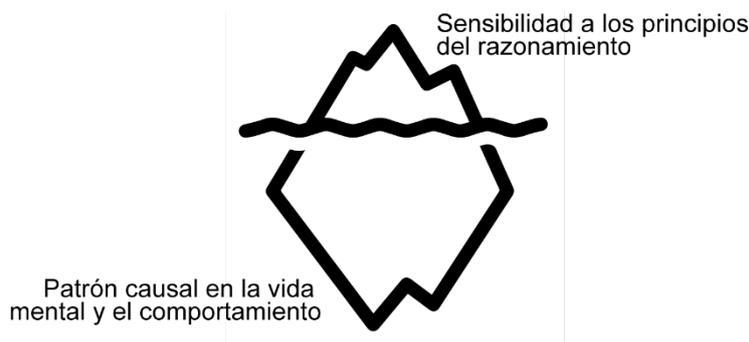
Empiezo por recordar que la creencia, a diferencia del juicio, es un estado permanente que afecta a largo plazo la vida mental y el comportamiento del sujeto. Por su misma naturaleza, una creencia introduce, a través de su contenido, un patrón causal tal que habrá un conjunto de fenómenos mentales y conductuales nuevos cuya explicación (o interpretación) requiere la atribución de esa creencia. En este sentido, el rol funcional de la creencia es almacenar un contenido para ser utilizado posteriormente por los distintos sistemas ‘consumidores’ de información, por decirlo en la jerga informacional. En el caso normal, la creencia no tiene una dinámica cognitiva independiente sino que está bajo el control directo del sujeto, lo que significa que su patrón causal característico es sensible a los principios que gobiernan el razonamiento teórico y práctico. Esto implica más específicamente que, en el caso normal, puede modificarse una creencia, o adquirir una nueva,

parece que, al menos en una primera fase, no estará dispuesto a aseverar la oración de la forma “ p ” (Moran 2001; Finkelstein 2003). Lo mismo puede pasarle al profesor cuando, a partir de ciertos indicios de su comportamiento discriminatorio, llegue a formarse la creencia de que él cree que no p : en una primera fase estará dispuesto a aseverar la oración de la forma “Creo que no p ” sin estar dispuesto a aseverar la oración de la forma “No p ”.

mediante la deliberación racional. En la medida en que el patrón causal característico de la creencia es sensible a los principios que gobiernan el razonamiento, podrá producirse una modificación doxástica a través de un procesamiento de arriba-abajo: la deliberación racional termina en un juicio que lleva a formarse una nueva creencia y, por lo tanto, a introducir un nuevo patrón causal en la vida mental y el comportamiento. Ahora bien, puede definirse una creencia *inconsciente* (dejando de lado si se trata del inconsciente freudiano o de otro tipo de inconsciente)⁷ como aquella creencia que tiene una dinámica cognitiva independiente. En particular, una creencia inconsciente es aquella que tiene una dinámica o rol funcional que no depende de tener la correspondiente creencia de segundo orden (Shoemaker 2009, 40-43; Vidal 2014 y 2018), lo que puede explicar que no esté bajo el control directo del sujeto. Esa dinámica cognitiva independiente puede estar determinada por un componente emocional del estado, y en este sentido podría sugerirse que no se trata de una creencia sino de una cuasi-creencia [*alief*], a saber: un estado consistente en un hábito o inclinación a responder ante un estímulo manifiesto de una cierta manera (Gendler 2008, 557) que está en tensión con alguna creencia explícita del sujeto. Por ejemplo, si estoy en un puente colgante suspendido a 300 metros de altura, y el puente está hecho de un material transparente, puedo tener la creencia de que el puente es seguro, pero también tener la cuasi-creencia de que voy a caerme⁸. Sin embargo, según la definición dada antes es constitutivo de tener una creencia inconsciente que el sujeto no crea que él tiene esa creencia (a no ser que intervenga algún proceso posterior de interpretación), lo que, como muestra el ejemplo anterior, no es constitutivo de tener una cuasi-creencia (ya que suelo creer, o saber de algún modo inmediato, que tengo la cuasi-creencia de que voy a caerme). Pues bien, en los casos de creencia inconsciente resultará que la mera deliberación racional no es suficiente para modificar esa creencia sino que, por el contrario, debe intervenir un procesamiento de abajo-arriba que primero actúe sobre el patrón causal previamente consolidado. En otras palabras, el sujeto no podrá formarse una nueva creencia, una creencia consciente, mediante el razonamiento, sin producirse antes ciertos cambios cognitivos y conductuales a niveles más profundos. Será cuestión de modificar en primer lugar las reacciones asociativas espontáneas, por ejemplo, reaccionar con temor ante alguien de otra raza, que emanan de nuestra segunda naturaleza, en este caso de esa parte conformada por el inconsciente afectivo-evaluativo que precede a toda evaluación racional. En términos experimentales, se ha mostrado que el condicionamiento evaluativo es exitoso en eliminar, o al menos reducir, las asociaciones prejuiciosas comunes (Olson y Fazio 2006). Puesto que hablamos de niveles más o menos profundos, los efectos permanentes de una creencia pueden representarse como un iceberg cuya punta es la sensibilidad a los principios que gobiernan el razonamiento:

⁷ Recientemente se ha empezado a hablar del inconsciente *adaptativo*, un aspecto del cual es una especie de sistema inmune psicológico que nos protege de amenazas a nuestro equilibrio mental, y otro aspecto es la existencia de fenómenos como prejuicios o sesgos implícitos (Green 2018, cap. 5).

⁸ Debo esta comparación con la noción de cuasi-creencia al comentario de un árbitro de la revista.



Consideremos entonces qué ocurre con el profesor que tiene la creencia inconsciente de que los estudiantes extranjeros no son tan buenos como los nacionales. Tener la creencia inconsciente de que no p no solo explicará su comportamiento discriminatorio sino también el hecho de que no llegue a formarse la creencia de que p después de una deliberación racional. Pues, supongamos que dispone de fundadas razones para juzgar que p . Mi idea es que, a pesar de ello, no llegará a formarse la creencia de que p porque, en la medida en que la creencia de que no p no es sensible a los principios que gobiernan el razonamiento, no puede haber una modificación doxástica sin que, independientemente de la deliberación racional, se altere el patrón causal correspondiente a la creencia de que no p . Quizá si empieza a detectar indicios de su comportamiento discriminatorio y a corregirlo sobre la base del juicio resultante de la deliberación racional, el sujeto pueda crear las condiciones cognitivas y conductuales, como la reducción de asociaciones prejuiciosas, para una modificación de la creencia. Adviértase que este es un caso distinto de aquel en el que un sujeto tiene una creencia recalcitrante, que sería cuando una creencia *consciente* ha calado tan profundo en la vida mental y el comportamiento que el sujeto mantiene esa creencia a pesar de llegar a saber que no hay ninguna razón o evidencia para creer (Cassam 2014, 20-24)⁹.

Una objeción inmediata a este planteamiento apelaría a la distinción entre una modificación doxástica y la adquisición de una nueva creencia. Ciertamente, diría el objetor, es muy posible que no pueda modificarse la creencia inconsciente de que no p sin un procesamiento de abajo-arriba, pero de ahí no se sigue que el juicio de que p no lleve al sujeto a formarse la creencia de que p . Aunque la mera deliberación racional no pueda modificar la creencia de que no p , esto sería compatible con el hecho de que el sujeto llegue a adquirir la creencia de que p y de este modo llegue a tener creencias inconsistentes, como sostienen Coliva y Bilgrami. Pero ¿qué mostraría que el sujeto adquirió la nueva creencia? Aparte de referirse a la disposición a volver a juzgar ese contenido o a aseverar sinceramente la oración “Los estudiantes extranjeros son tan buenos como los nacio-

⁹ Es discutible que en realidad haya creencias recalcitrantes, puesto que se trataría de creencias conscientes que no son sensibles a los principios que gobiernan el razonamiento. Como es claro, mi propio punto de vista en este artículo es incompatible con su existencia.

nales”, algo que ya examinamos con anterioridad, el objetor podría referirse también a la disposición a utilizar ese contenido como una premisa en el razonamiento teórico y práctico. Ahora bien, téngase en cuenta, en primer lugar, que tales cambios cognitivos y conductuales deben ser relativamente superficiales, puesto que el caso del profesor está construido sobre el supuesto de que, después de su deliberación racional, él sigue teniendo un comportamiento discriminatorio generalizado. Parece así que la creencia de que p , que debería ser una creencia consciente, no habría introducido un patrón causal que, a niveles más profundos, sea sensible a los principios que gobiernan el razonamiento teórico y práctico, lo que significa que en realidad no puede hablarse de creencia en este caso. Evidentemente, la causa de ello es la persistencia de la creencia inconsciente de que no p con su patrón causal característico a esos niveles más profundos. Resulta, en segundo lugar, que hay una explicación más económica de los cambios cognitivos y conductuales al nivel del razonamiento. Una vez que el sujeto llega a juzgar que p , empezará a razonar *como si* tuviera la creencia de que p , lo que precisamente consiste en utilizar ese contenido como una premisa en el razonamiento teórico y práctico. Esto no implica que, en vez de tener la creencia de que p , el sujeto se encuentre en algún otro estado capaz de producir ciertos efectos propios de la creencia, como podría ser el estado de *simular* internamente (o imaginar o hacer-creer a uno mismo) que p es el caso según una explicación reciente del autoengaño (Gendler 2007)¹⁰. Tampoco es necesario sostener que, partiendo del juicio, el sujeto se encuentra en un estado de creencia caracterizado por la vaguedad de ser un estado entre [*in-between*] creer y no creer, en el sentido de que ni es correcto decir que el sujeto tiene la creencia de que p ni es correcto decir que no tiene la creencia de que p (Schwitzgebel 2010). Veremos que, para dar cuenta de los efectos permanentes al nivel del razonamiento, es suficiente apelar a la creencia de segundo orden que llegó a formarse a partir del juicio. Cuando más adelante el sujeto advierta indicios de su comportamiento discriminatorio e intente corregirlo, se producirán cambios cognitivos y conductuales a niveles cada vez más profundos, los que, si llegaran a alcanzar el inconsciente afectivo-evaluativo, podrían llevar a la adquisición de la creencia de que p . En ese momento, el patrón causal característico de la creencia de que no p habrá sido reemplazado por el patrón causal característico de la creencia de que p .

Como hemos visto, parece que, aunque no llegue a formarse la creencia de primer orden, el profesor llega a creer que él cree que los estudiantes extranjeros son tan buenos como los nacionales. Esto es lo que explicaría que esté dispuesto a aseverar sinceramente la correspondiente oración de la forma “Creo que p ” (y así, por equivalencia performativa, la correspondiente oración de la forma “ p ”). Pero ¿cómo ocurre que la creencia inconsciente de que no p que ya tenía afectara el proceso de formación de la creencia de que p basado en la deliberación racional y, sin embargo, el sujeto pudo llegar a formarse la creencia de que él cree que p ? Mi punto es considerar la existencia de una conexión

¹⁰ Desde luego, en el planteamiento de Gendler el sujeto tampoco llegó a juzgar que p para luego simular la creencia, en vez de creer, sino que simular internamente que p es el caso es su actitud desde el principio.

más lábil entre las creencias de segundo orden y la conducta, en un sentido amplio que comprende tanto la vida mental como el comportamiento, que la que existe entre las creencias de primer orden y esa misma conducta. Ciertamente, en condiciones normales, tanto la creencia de segundo orden como la creencia de primer orden llevan al sujeto a conducirse sobre la base de que p es un contenido verdadero. Este es el caso en el que el sujeto ejerce la agencia racional o autoría (Boyle 2011; Moran 2001 y 2012; Parrott 2017), cuando la creencia de segundo orden se forma en el proceso de autodeterminarse o configurar la propia mente mediante la deliberación racional (es decir, cuando el juicio de que p lleva a formarse activamente la creencia de que p)¹¹. Pues, en ese caso tener la creencia de segundo orden es otro modo de *endosar* el contenido de la creencia de primer orden como parte de su perspectiva sobre el mundo. Pero una creencia de segundo orden no introduce, por su misma naturaleza, el patrón causal característico de la creencia de que p . De otro modo, no podríamos tener creencias de segundo orden como resultado de adoptar una actitud teórica (u observacional) hacia nuestras creencias de primer orden, en el sentido preciso de adoptar la misma actitud que adoptamos hacia las creencias de los demás como meros hechos psicológicos, como ocurre, por ejemplo, en un proceso de interpretación psicoanalítica. Como es obvio, la creencia de que alguien más cree que p , que también es una creencia de segundo orden, no introduce en *nosotros* el patrón causal correspondiente a la creencia de que p .

He argumentado que, después de su deliberación racional, el profesor no llega a formarse la creencia de que p porque la creencia de que no p que ya tenía no es sensible a los principios que gobiernan el razonamiento. Puesto que, por su misma naturaleza, la creencia de que p introduce un patrón causal característico, no puede adquirirse esa creencia sin alterar, por otros medios que la deliberación racional, el patrón causal característico de la creencia de que no p . Veámoslo de modo inverso. Si, contrariamente a su naturaleza, pudiera adquirirse la creencia de que p sin introducir un patrón causal característico, entonces tener la creencia inconsciente de que no p , con su propio patrón causal no sensible al razonamiento, no impediría que el sujeto se forme la creencia de que p a través de la deliberación racional. En ese caso la creencia de que p solo tendría efectos al nivel del razonamiento. Justamente acabo de defender la posibilidad de formarse una creencia de segundo orden sin introducir el patrón causal correspondiente a la creencia de primer orden. La conexión lábil entre la creencia de segundo orden y el patrón causal

¹¹ Boyle piensa en términos de un autoconocimiento tácito que solo posteriormente se manifestaría a la reflexión, lo que no excluye que ese autoconocimiento consista en tener una creencia tácita de segundo orden. Estoy considerando fundamentalmente su versión kantiana de la autoría como una actividad de la razón, aunque la propuesta original es de Moran. Para Parrott, sin autoconocimiento genuino no tendría sentido la sensibilidad de las creencias al juicio entendida como la conexión racional inmediata entre el juicio acerca de lo que uno debe creer y las creencias que de hecho uno tiene. Más recientemente, Moran ha desarrollado también la idea de que, puesto que un sujeto es racionalmente responsable de lo que hace, la pregunta por las razones de la acción, entre ellas la creencia, solo es inteligible como una pregunta que el propio sujeto puede responder, lo que implica tener autoconocimiento genuino.

característico de la creencia de que p explica así que pueda adquirirse esa creencia sin alterar el patrón causal característico de la creencia de que no p . Esta es la explicación de que tener previamente la creencia inconsciente de que no p no afecte la formación de la creencia de segundo orden como una creencia que puede tener los mismos efectos que la creencia de que p al nivel del razonamiento. Obtenemos así el siguiente cuadro sobre los efectos permanentes, respecto a un cierto contenido p , de los distintos tipos de creencia en términos de su naturaleza cognitiva:

Tipos de creencia	Sensibilidad a los principios del razonamiento	Patrón causal en la vida mental y comportamiento
Creencia consciente de que p	SI	SI
Creencia inconsciente de que p	NO	SI
Creencia de que uno cree que p	SI	NO

Pero, ¿cuál es el proceso por el que el sujeto llegaría realmente a formarse una creencia de segundo orden? En el caso del profesor, es claro que la creencia de que él cree que p no se forma como resultado de adoptar una actitud teórica (u observacional). Pero esa creencia de segundo orden tampoco se forma en el proceso de configurar la propia mente a través de la deliberación racional, en la medida en que el sujeto no llega realmente a configurarla (es decir, no llega a formarse activamente la creencia de que p). No puede decirse, por lo tanto, que este sea un caso en el que tener la creencia de segundo orden es endosar el contenido de la creencia de primer orden (y por ello conducirse de igual manera a todos los niveles). Nos encontramos entonces, respecto a la formación de la creencia de segundo orden, ante un caso intermedio entre la actitud teórica y la actitud propiamente autodeterminante. Me parece que el caso puede describirse del siguiente modo. El profesor llega a juzgar racionalmente que p . Tener ya la creencia inconsciente de que no p le impide pasar del juicio de que p a la creencia de que p , dada la naturaleza cognitiva de las creencias de primer orden. Pero no le impide formarse la creencia de que él cree que p , dada la naturaleza cognitiva de las creencias de segundo orden. Ahora bien, ha estado involucrado en el mismo tipo de deliberación que, en los casos de genuina agencia racional, le lleva a formarse una creencia de segundo orden en el proceso de configurar su propia mente. Se forma entonces la creencia de que él cree que p . Sin embargo, no ha (re)configurado realmente su mente, por lo que no llega a introducir en su vida mental y su comportamiento el patrón causal característico de la creencia de que p .

Termino con dos consideraciones sobre esta cuestión. Primero, téngase en cuenta que formarse esa creencia de segundo orden puede explicar, no solo que el profesor esté dispuesto a aseverar sinceramente la correspondiente oración de la forma “Creo que p ” (y así de la forma “ p ”), sino también a utilizar el contenido p como una premisa en el razonamiento teórico y práctico. Pues, aunque lábil, es muy plausible que, como sugerí antes, siga habiendo una conexión entre la creencia de segundo orden y la conducta (en

sentido amplio) propia de la creencia de que p al nivel del razonamiento, a diferencia de lo que ocurre cuando la formación de esa creencia es el resultado de adoptar una actitud meramente teórica. Al fin y al cabo esa creencia se formó como resultado de un proceso de deliberación racional que termina en el juicio de que p . Con todo, el caso descrito es tal que la creencia de que él cree que p no es capaz de introducir un patrón causal, el patrón causal característico de la creencia de que p , a niveles más profundos. Segundo, esto permite una reflexión sobre lo que está mal con el constitutivismo. Recordemos que, para el constitutivista, es una verdad o necesidad conceptual que si, en condiciones normales, un sujeto cree que él cree que p , entonces él cree que p . En la versión de Wright, esas condiciones normales consistían básicamente en la lucidez y la posesión de los conceptos relevantes (el concepto de creencia y el concepto de primera persona) en la formación de la creencia de segundo orden. El problema es que esta caracterización de las condiciones normales parece insuficiente. De entrada, es una caracterización que no distingue entre formarse la creencia de segundo orden como resultado de adoptar una actitud teórica y como resultado de adoptar una actitud deliberativa, dado que en ambos casos debe haber lucidez cognitiva y posesión de los conceptos relevantes. Pero en el caso de adoptar una actitud meramente teórica en un proceso de interpretación o de observación de la propia conducta, es claro que la probabilidad del error es tal que no es una verdad conceptual que si el sujeto tiene la creencia de segundo orden, entonces tiene la correspondiente creencia de primer orden. A este respecto, Coliva y Bilgrami rectifican la posición de Wright incluyendo entre las condiciones normales el requerimiento de que el sujeto esté involucrado en una deliberación racional (que le llevaría a formarse activamente la creencia de que p como un compromiso). El problema ahora es que, según el análisis desarrollado aquí, esta sigue siendo una caracterización insuficiente. Como hemos visto, la deliberación racional no garantiza que lleguemos a configurar nuestra mente en términos de la creencia de que p . Una caracterización adecuada de las condiciones normales requeriría mencionar también el hecho de que el sujeto no tiene la creencia inconsciente de que no p . Sin embargo, esta formulación se vuelve así susceptible a la acusación de ser una formulación *ad hoc*, que es justamente la acusación que Coliva dirige a Wright cuando él, en algún momento, incluye entre las condiciones normales la condición de que el sujeto no haya caído en el autoengaño (Coliva 2016, 198).

5. Autoconocimiento, autoengaño y transparencia

Pasemos entonces a examinar el modo en que el análisis propuesto aquí resulta crítico para la teoría de la transparencia de Byrne. Recordemos que, según Byrne, el carácter peculiar y privilegiado del autoconocimiento de las creencias puede explicarse en términos del seguimiento de la regla doxástica:

(CRE) Si p , cree que tú crees que p .

Como vimos, Byrne sostiene, frente a algunos objetores, que puede dar cuenta así tanto del autoconocimiento genuino de una creencia consciente que el sujeto ya tenía antes de seguir esa regla como del autoconocimiento genuino de la creencia que llega a formarse cuando delibera ahora sobre si p es o no es el caso. Pero, en general, esos mismos objetores no cuestionan la teoría de Byrne respecto al autoconocimiento de las creencias basado en la deliberación racional, lo que significa que su crítica solo afecta al alcance explicativo de la teoría. Mi objeción es más relevante en la medida en que está dirigida contra el poder explicativo de la teoría también allí donde se suponía que era más fuerte. (No obstante, me ocuparé también del autoconocimiento genuino de creencias conscientes previas). Byrne argumenta que las creencias de segundo orden que un sujeto se forma siguiendo la regla (CRE) son autoverificantes porque, para poder seguir esa regla¹², el sujeto debe tener la creencia de que p (algo que él no parece separar de haber juzgado que p en la medida en que normalmente un juicio resulta en un estado de creencia). Pero el caso del profesor muestra que no es así cuando el sujeto ya tenía la creencia inconsciente de que no p . En efecto, supongamos primero que el profesor está involucrado ahora en una deliberación racional, en una evaluación de la evidencia disponible, sobre si p es o no es el caso. A partir de fundadas razones el sujeto puede llegar a juzgar que p sin formarse también la creencia de que p como un estado permanente, dado que, según argumenté, la deliberación racional no puede introducir en este caso el patrón causal correspondiente a la creencia de que p . Esto no es óbice para que el procedimiento por el cual el sujeto llega a formarse la creencia de segundo orden es seguir la regla (CRE). Pues, es claro que puede seguir esa regla si ha llegado a juzgar que p como resultado de su deliberación racional. Téngase en cuenta que, en mi planteamiento, no está en cuestión que el profesor se forme la creencia de que él cree que p , lo que puede explicarse por la conexión lábil entre las creencias de segundo orden y la conducta (en sentido amplio). Se trataría entonces, concediendo el punto a Byrne, de que el procedimiento empleado por el sujeto para formarse la creencia de segundo orden es el seguimiento de la regla doxástica. El problema es que esa creencia de segundo orden no es autoverificante, puesto que el sujeto no ha llegado a formarse la creencia de que p . Seguir la regla (CRE) no es, por lo tanto, una condición suficiente para tener un autoconocimiento genuino de las creencias (ni siquiera bajo la restricción más favorable a los casos donde interviene la deliberación racional).

En defensa de la teoría de Byrne podría apuntarse que las condiciones normales para el seguimiento de la regla (CRE) deben incluir el hecho de que el sujeto no tiene la creencia inconsciente de que no p . Pero, como ocurría con la teoría constitutivista de Wright, esta parece ser una solución *ad hoc*, especialmente considerando que los casos de autoengaño, sobre todo los casos de sesgo implícito, pueden estar mucho más extendidos de lo que solemos pensar (Staats 2017). Más aún, la teoría de Byrne comparte con las teorías racionalistas del autoconocimiento (Shoemaker 1996) la idea de que para tener autocono-

¹² Por simplicidad, no tendré en consideración la circunstancia en la que p no es el caso y, por lo tanto, el sujeto no puede seguir la regla doxástica sino solamente tratar de seguirla, como vimos en la sección 2.

cimiento genuino no se requiere una facultad dedicada exclusivamente al conocimiento de la propia mente, como un sentido interno, sino estar en posesión de las mismas capacidades epistémicas que se emplean en otros dominios, como la racionalidad inferencial y la capacidad conceptual (Byrne 2018, 14, 112-114). De este modo, las condiciones normales que deben considerarse son aquellas en las que esas capacidades se ejercen habitualmente (lo que excluye, por ejemplo, estar drogado). Justamente el profesor no tiene ningún impedimento en cuanto a su capacidad de razonar inferencialmente cuando sigue la regla (CRE), ni en cuanto a su capacidad conceptual, la posesión de los conceptos de creencia y de primera persona, en la formación de la creencia de segundo orden. En consecuencia, el seguimiento de la regla doxástica debería proporcionar autoconocimiento, que es lo que no ocurre en este caso.

Supongamos ahora que el profesor no está involucrado en una deliberación racional sobre si p es o no es el caso, pero llega a juzgar que p . Se podría pensar que el juicio es el resultado de tener previamente la creencia consciente de que p , lo que implica que este contenido fue retenido para utilizarlo posteriormente, como ocurre ahora en el juicio (y de ahí seguir la regla doxástica para formarse la creencia de segundo orden). Pero ya sabemos que esta no es una opción viable, dado que el sujeto tenía la creencia inconsciente de que no p . Supongamos entonces que en algún momento del pasado llegó a formarse la creencia de que él cree que p (sin tener realmente la creencia de que p). Una explicación alternativa del juicio apelaría a esa creencia de segundo orden que el sujeto ya tenía previamente, en la medida en que, según he argumentado, esa creencia puede llevar a que razone *como si* creyera que p , lo que incluye realizar el juicio de que p . Por más que uno esté tentado a concluir que así no habría una diferencia real con el caso en el que un sujeto ya tenía la creencia consciente de que p , debe recordarse que no puede haber creencia sin un patrón causal característico que no solo se manifieste al nivel del juicio (o del razonamiento) sino también a niveles más profundos de la vida mental y del comportamiento. Precisamente el profesor no estaba respondiendo, a esos niveles más profundos, en términos del patrón causal característico de la creencia de que p sino en términos del patrón causal característico de la creencia de que no p . Ahora bien, en esta versión del caso del profesor, resulta que él ni siquiera estaría en condiciones de seguir la regla (CRE). No tiene sentido que, después de juzgar que p , siga la regla doxástica para formarse la creencia de que él cree que p , puesto que la realización misma del juicio depende precisamente del hecho de que el sujeto *ya* tenía esa creencia de segundo orden. Podemos concluir entonces que la teoría de la transparencia de Byrne no establece condiciones suficientes para tener un autoconocimiento genuino ni en los casos en los que el sujeto está involucrado en una deliberación racional ni, desde luego, en los casos en los que el sujeto llega a realizar un juicio sin deliberación.

6. Conclusión

En este artículo, he presentado una objeción a la explicación del autoconocimiento genuino de nuestras creencias en términos del método de la transparencia (y de la deliberación racional), considerando, en particular, la versión de Byrne. Para ello examiné un caso de autoengaño en el que un sujeto llega a realizar un determinado juicio, un juicio de que p , pero manifiesta un comportamiento discriminatorio incompatible con ese juicio. Argumenté, primero, que la interpretación más económica de este caso es aquella según la cual el sujeto tenía previamente la creencia inconsciente de que no p y, además, el juicio de que p , un acto mental ocurrente, no llega a transformarse en un estado de creencia. A continuación sostuve que, puesto que una creencia inconsciente es aquella que no es sensible a los principios que gobiernan el razonamiento teórico y práctico, tener la creencia inconsciente de que no p no solo explica el comportamiento discriminatorio sino también el hecho de que, a pesar de realizar el juicio de que p como resultado de una deliberación racional, el sujeto no llega a formarse la correspondiente creencia. Más concretamente, mi punto es que el comportamiento discriminatorio muestra que la deliberación racional no fue capaz de introducir en la vida mental y el comportamiento del sujeto el patrón causal característico de la creencia de que p , lo que se debería a que él tiene la creencia inconsciente de que no p . Sin embargo, en la medida en que la conexión entre la conducta (en sentido amplio) y las creencias de segundo orden es más lábil, no parece que haya un impedimento a que el sujeto se forme la creencia de que él cree que p como resultado de estar involucrado en el mismo proceso, la deliberación racional, por el que se forma esa creencia de segundo orden cuando realmente logra configurar su mente (es decir, cuando el juicio de que p se transforma en la creencia de que p). Ocurre entonces que en el caso de autoengaño bajo consideración un sujeto llega a formarse una creencia de segundo orden sin haber adquirido la correspondiente creencia de primer orden, lo que obviamente significa que no hay autoconocimiento ni, mucho menos, autoconocimiento genuino. Ahora bien, la teoría de Byrne propone, como método de la transparencia, el seguimiento de una regla doxástica que debería producir autoconocimiento genuino cuando el sujeto está involucrado en una deliberación racional. La conclusión es que la teoría de la transparencia de Byrne no establece condiciones suficientes para tener un autoconocimiento genuino de las creencias. (Adicionalmente, argumenté que su teoría tampoco funciona cuando se trata del autoconocimiento genuino de creencias conscientes que uno ya tenía).

Agradecimientos

Este artículo fue escrito gracias al apoyo financiero otorgado por el gobierno de Chile mediante el proyecto Fondecyt n° 1170039.

Referencias bibliográficas

- Bar-On, D. (2004). *Speaking My Mind*. Oxford: Oxford University Press. doi:10.1093/0199276285.001.0001
- Bilgrami, A. (2006). *Self-Knowledge and Resentment*. Cambridge Mass.: Harvard University Press.
- Boyle, M. (2011). Transparent Self-Knowledge. *Proceedings of the Aristotelian Society*, Suplementario 85(1): 223-241. doi:10.1111/j.1467-8349.2011.00204.x
- Byrne, A. (2018). *Transparency and Self-Knowledge*. Oxford: Oxford University Press. doi:10.1093/oso/9780198821618.001.0001
- Cassam, Q. (2014). *Self-Knowledge for Humans*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199657575.001.0001
- Coliva, A. (2016). *The Varieties of Self-Knowledge*. Copenhagen: Palgrave MacMillan. doi:10.1590/0100-6045.2017.v40n3.ab
- Evans, G. (1982). *The Varieties of Reference*. Oxford: Clarendon Press.
- Fernández, J. (2013). *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199664023.001.0001
- Fernández Acevedo, G. (2018). *El autoengaño: Anatomía de una pasión humana*. Buenos Aires: SADAF.
- Finkelstein, D. (2003). *Expression and the Inner*. Cambridge Mass.: Harvard University Press.
- Fricke, M. (2019). Estrategias para solucionar el problema del autoconocimiento. En J. Vidal y C. Muñoz (coord.), *Perspectivas sobre la subjetividad*, pp. 181-222. Concepción: Editorial de la UdeC.
- Gallois, A. (1996). *The World Without, The Mind Within*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511627880
- Gendler, T. S. (2008). Alief in Action (and Reaction). *Mind and Language*, 23: 552-585. doi:10.1111/j.1468-0017.2008.00352.x
- Gendler, T. S. (2007). Self-Deception as Pretense. *Philosophical Perspectives*, 21: 231-258. doi:10.1111/j.1520-8583.2007.00127.x
- Gertler, B. (2011). Self-Knowledge and the Transparency of Belief. En A. Hatzimoysis (ed.), *Self-Knowledge*, pp. 125-145. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199590728.003.0007
- Green, M. S. (2018). *Know Thyself: The Value and Limits of Self-Knowledge*. New York: Routledge. doi:10.4324/9781315560298
- Mele, A. (2000). *Self-Deception Unmasked*. Princeton: Princeton University Press. doi:10.1515/9781400823970



- Moran, R. (2001). *Authority and Estrangement*. Princeton: Princeton University Press. doi:10.1515/9781400842971
- Moran, R. (2012). Self-Knowledge, “Transparency”, and the Forms of Activity. En D. Smithies y D. Stoljar (eds.), *Introspection and Consciousness*, pp. 211-236. New York: Oxford University Press. doi:10.1093/acprof:oso/9780199744794.003.0008
- Olson, M. A., Fazio, R. H. (2006). Reducing Automatically Activated Racial Prejudice through Implicit Evaluative Conditioning. *Personality and Social Psychology Bulletin*, 32(4): 421-433. doi:10.1177/0146167205284004
- Parrott, M. (2017). Self-Blindness and Self-Knowledge. *Philosophers’ Imprint*, 16: 1-22. <http://hdl.handle.net/2027/spo.3521354.0017.016>
- Peacocke, C. (1999). *Being Known*. Oxford: Oxford University Press. doi:10.1093/0198238606.001.0001
- Rosenthal, D. (2005). *Consciousness and Mind*. New York: Oxford University Press.
- Schwitzgebel, E. (2010). Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly*, 91: 531-553.
- Shoemaker, S. (2009). Self-Intimation and Second-Order Belief. *Erkenntnis*, 71(1): 35-51. doi:10.1007/s10670-009-9172-z
- Shoemaker, S. (1996). Self-Knowledge and “Inner Sense” II. En S. Shoemaker, *The First-Person Perspective and Other Essays* pp. 224-245. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511624674.009
- Staats, C. (2017). State of the Science: Implicit Bias Review 2017. Ohio: Kirwan Institute, State University. <http://kirwaninstitute.osu.edu/implicit-bias-training/resources/2017-implicit-bias-review.pdf>
- Vidal, J. (2014). Creencia de primera persona, conciencia y la paradoja de Eroom. *Crítica*, 46(138): 37-64.
- Vidal, J. (2018). Respuesta a Sebastián: en defensa de la naturaleza consciente de las creencias de primera persona. *Crítica*, 50(150): 65-89. doi:10.22201/iifs.18704905e.2018.16
- Wright, C. (2001). *Rails to Infinity*. Cambridge Mass.: Harvard University Press.