Tópicos (México)
ISSN: 0188-6649
Universidad Panamericana, Facultad de Filosofía

Fricke, Martín Francisco
¿BEL o Bypass? Dos teorías de la transparencia del autoconocimiento
Tópicos (México), núm. 59, 2020, Julio-Diciembre, pp. 11-50
Universidad Panamericana, Facultad de Filosofía

DOI: https://doi.org/10.21555/top.v0i59.1101

Disponible en: https://www.redalyc.org/articulo.oa?id=323064336001

http://doi.org/10.21555/top.v0i59.1101

BEL or Bypass?

Two Transparency Theories of Self-Knowledge

¿BEL o Bypass?

Dos teorías de la transparencia del autoconocimiento

Martín Francisco Fricke
Instituto de Investigaciones Filosóficas,
Universidad Nacional Autónoma de México, UNAM
Escuela Nacional de Estudios Superiores, Unidad Mérida,
Universidad Nacional Autónoma de México, UNAM
México
mfcephcis@gmail.com

Recibido: 10 - 10 - 2018.
Aceptado: 16 - 01 - 2019.

Abstract

Alex Byrne and Jordi Fernández propose two different versions of a transparency theory of self-knowledge. According to Byrne, we self-attribute beliefs by an inference from what we take to be facts about the world (following a rule he calls Bel). According to Fernández, we self-attribute the belief that *p* on the basis of a prior mental state, a state which constitutes our grounds for the belief that *p* (thereby realizing a procedure he calls Bypass). In this paper, I present the two theories in outline and discuss various objections concerning their normative (Can the procedure give us *knowledge*?) and metaphysical aspects (Is the procedure *functional*?). I conclude that especially the metaphysical objections against Bypass are somewhat more difficult to counter than those against Bel and that the modifications required of Fernández's theory make it very similar to Byrne's.

Keywords: self-knowledge; transparency theories; Alex Byrne; Jordi Fernández; Gareth Evans.

Resumen

Alex Byrne y Jordi Fernández proponen dos diferentes versiones de la teoría de la transparencia del autoconocimiento. Según Byrne, para autoatribuir creencias inferimos qué es lo que creemos a partir lo que tomamos como hechos sobre el mundo (siguiendo una regla que Byrne llama Bel). Según Fernández, autoatribuimos la creencia de que *p* con base en un estado anterior a esta creencia, un estado que fundamenta la creencia de que *p* (realizando un procedimiento que él llama Bypass). En este artículo expongo las dos teorías y discuto objeciones que conciernen su aspecto normativo (¿puede el procedimiento darnos *conocimiento*?) y metafísico (¿es *funcional* el procedimiento?). Concluyo que en especial las objeciones metafísicas son más graves en el caso de Bypass que en el de Bel y que las modificaciones requeridas de la teoría de Fernández la asemejan mucho a la de Byrne.

Palabras clave: autoconocimiento; teorías de la transparencia; Alex Byrne; Jordi Fernández; Gareth Evans.

1. Introducción

En este artículo discuto dos teorías del autoconocimiento: la de Alex Byrne y la de Jordi Fernández. Ambas pueden ser clasificadas como teorías de la transparencia del autoconocimiento (*transparency theories*), pero las dos varían en algunos detalles importantes. Mi propósito aquí es exponer las teorías, elaborar algunas objeciones centrales contra cada una y compararlas. Como se verá en la discusión, me parece que la teoría de Fernández se enfrenta a objeciones más graves que la de Byrne; en especial, no está claro si la teoría puede dar cuenta del conocimiento que tenemos de nuestras creencias si las formamos de manera accidental o si nuestras maneras de formarlas cambian. Si tratamos de modificar la propuesta de Fernández en respuesta a estas objeciones, observaremos que las diferencias entre Fernández y Byrne disminuyen y tal vez desaparecen. Espero que mi discusión muestre tanto la promesa de las teorías de la transparencia para explicar el autoconocimiento como la importancia de responder a algunas objeciones centrales en contra de ellas.

Para entrar en la discusión es conveniente primero aclarar cuál es el problema que Byrne y Fernández tratan de solucionar: el problema del autoconocimiento (sección 2), y elucidar, en términos generales, la noción de transparencia (sección 3). Después expongo cada teoría y discuto dos objeciones en contra de cada una (secciones 4 a 7).[1]

2. El problema del autoconocimiento

El conocimiento que tenemos de algunos de nuestros estados mentales tiene dos características que están en tensión. Según Alex Byrne, podemos llamarlas "acceso privilegiado" y "acceso peculiar" (cfr. Byrne, 2005, pp. 80-82).[2] El *acceso privilegiado* consiste en el hecho de que existe

[1] Las secciones 4 (Byrne), 5 (objeciones a Byrne) y 6 (Fernández) exponen principalmente posiciones de otros autores; por otra parte, las demás secciones y las respuestas a las objeciones en la sección 5, en la mayor parte, desarrollan ideas mías.

[2] Aunque mi caracterización de los dos tipos de acceso sigue la de Byrne, la descripción del conflicto entre los dos no es de Byrne, sino mía. De hecho, ni Byrne ni Fernández hablan de un *conflicto* entre las dos características, sino sólo de la necesidad de explicar los dos tipos de acceso.

poca probabilidad de error cuando adquirimos el autoconocimiento en cuestión. Nuestras creencias sobre los estados mentales aquí bajo consideración tienen una muy alta probabilidad de ser conocimientos. Cuando creo: "Creo que X ganará las elecciones", es posible que me equivoque con respecto al ganador de las elecciones. Pero hay muy poca probabilidad de que no sea cierto que *yo crea* que X ganará las elecciones. Algunos incluso afirmarían que es imposible equivocarse sobre este tipo de estados mentales cuando son los propios.

A diferencia del acceso privilegiado, el *acceso peculiar* en nuestro autoconocimiento no tiene que ver con una cuestión normativa (¿es mi creencia sobre algún estado mental mío un conocimiento?), sino con la manera en que adquirimos este conocimiento. Lo adquirimos de una manera que es peculiar por dos razones:[3] primero, no parece depender de observaciones ni de inferencias elaboradas. Segundo, es una manera de adquirir conocimientos que sólo funciona para uno mismo, para conocer los propios estados mentales (algunos de ellos, por lo menos), no para saber de los estados mentales de otras personas o para saber del resto del mundo. Cuando se trata de otras personas, tenemos que observarlas, tal vez conversar con ellas, consultarnos con otras personas y hacer inferencias para averiguar qué es lo que sucede en sus mentes. En cambio, cuando se trata de nosotros mismos parece que no necesitamos observaciones de nuestro propio comportamiento y tampoco necesitamos inferencias complicadas e inciertas para saber qué es lo que creemos. Algunos dirían que el conocimiento en cuestión no es inferencial, sino inmediato.[4]

El problema del autoconocimiento como yo lo entiendo consiste en la aparente incompatibilidad de acceso privilegiado y acceso peculiar. El acceso privilegiado, por sí solo, asociado a un conocimiento también requiere una explicación; pero por lo menos a primera vista tal explicación no es tan difícil de dar. Asimismo, el acceso peculiar, por sí solo, no parece tan problemático. El problema del autoconocimiento es que comprende *ambas* características *al mismo tiempo*: el acceso privilegiado *y* el acceso

[3] Fernández enfatiza la primera razón; Byrne, la segunda. Fernández utiliza el término "acceso especial" para hablar de lo que aquí llamo, con Byrne, "acceso peculiar".

[4] Cfr. Tugendhat (1993), por ejemplo. Como veremos más adelante, a diferencia de Fernández, Byrne piensa que el autoconocimiento sí es inferencial; pero las inferencias necesarias, según él, son extremadamente simples.

peculiar. ¿Por qué parecen estar en conflicto las dos características? La razón es que lo que ordinariamente confiere privilegio a algún acceso epistémico consiste en factores que están ausentes si el acceso también es peculiar en el sentido descrito anteriormente.

Veamos un caso que no tiene que ver con el autoconocimiento, para distinguir lo que ordinariamente confiere privilegio a un acceso epistémico. Un doctor tiene un acceso privilegiado al estado de salud de su paciente, por ejemplo, a través de placas de rayos X. ¿Por qué su acceso epistémico es privilegiado? Porque su percepción es especialmente entrenada, porque tiene instrumentos especiales de investigación (el aparato de rayos X), porque conoce las relevantes teorías médicas y porque tal vez se consulta con otros especialistas en el área. ¿Es su acceso al estado de salud de su paciente también peculiar? Para ser peculiar no debería depender de observaciones o inferencias elaboradas y sólo debería ser disponible para una sola persona. Pero eso es claramente incompatible con la idea de que el doctor tenga un acceso privilegiado a la salud de su paciente. El doctor no puede conocer bien el estado del paciente si no lo observa repetidamente y usando métodos e instrumentos especiales. Tampoco podemos confiar en sus opiniones si no los puede respaldar con teorías médicas aceptadas y si se rehúsa a escuchar lo que otros especialistas opinen. Si mantiene que *sólo él* puede, en principio, conocer el estado de salud de su paciente, está claro que deberíamos buscarnos otro médico. En resumen, cuando se trata del acceso epistémico que un doctor tiene al estado de salud de su paciente, privilegio es incompatible con peculiaridad en los sentidos anteriormente especificados. Si el acceso es privilegiado no puede ser peculiar, y si es peculiar no puede ser privilegiado.

Este punto puede generalizarse. Ordinariamente, para hacer nuestros intentos de conocimiento menos propensos a ser erróneos y en este sentido más privilegiados, tenemos que recurrir a métodos que hacen que el conocimiento *no* sea peculiar. Tenemos que hacer observaciones adicionales, usar instrumentos especiales, respaldarnos con teorías ya probadas o consultarnos con otros especialistas en el área. Todo esto significa que nuestro acceso al asunto en cuestión *no* será peculiar, porque no será independiente de la observación y de inferencias elaboradas y no será exclusivo, por su método, a una sola persona.

Lo que he elaborado en los dos párrafos anteriores también es correcto cuando se trata del conocimiento de los estados mentales de otras

personas. Es posible que alguien sea especialista para conocer la mente de otra persona: tal vez un psicólogo pueda tener un acceso privilegiado a la mente de un paciente. Pero como en el caso del radiólogo, su acceso no parece ser peculiar. Si el psicólogo conoce la mente de su paciente mejor que otros es porque lo ha observado repetidamente y durante un tiempo extendido, ha conversado con él, ha hecho diversas pruebas psicológicas, conoce teorías especializadas sobre personas como él y se ha consultado con otros psicólogos sobre el caso. Todo eso significa que su acceso a la mente de su paciente *no* es peculiar: no es muy directo, sí depende de observaciones y no es exclusivo de una sola persona, sino en principio abierto a cualquier otro psicólogo que quiera hacerse conocedor del caso. Tenemos, entonces, que el acceso privilegiado a la mente de otra persona no puede ser peculiar también.

Sin embargo, las cosas son diferentes cuando se trata de los *propios* estados mentales. El acceso que tenemos a ellos no depende de observaciones ni de inferencias complicadas o de consultas con otras personas y sólo es disponible para una persona. Y a pesar de esta peculiaridad, nuestro acceso a por lo menos algunos de los propios estados mentales es muy poco propenso al error, es decir, es privilegiado. En resumen, nuestro autoconocimiento exhibe al mismo tiempo privilegio y peculiaridad. El problema del autoconocimiento consiste en explicar cómo eso es posible. ¿Cómo es posible que tengamos privilegio, aunque, por la peculiaridad del autoconocimiento, no tenemos ninguno de los rasgos que normalmente lo confieren?

En lo que sigue examinaré dos teorías que tratan de solucionar el problema del autoconocimiento con referencia a la así llamada "transparencia de la mente". Durante todo el texto sólo hablaré del autoconocimiento de las propias *creencias*, aunque se supone y espera que la solución pueda ser generalizada al conocimiento privilegiado y peculiar de otras actitudes proposicionales también como son los deseos, intenciones, esperanzas, pensamientos etc.

3. ¿Qué significa "transparencia de la mente"?

Un material es transparente, en el sentido ordinario, si deja pasar la luz. El cristal de una ventana es transparente porque nos permite ver desde nuestro lado qué hay en el otro lado de la ventana. Mientras más transparente, menos visible es el material mismo. Medusas usan su transparencia como camuflaje: debido a ésta, difícilmente son visibles para otros animales.

En un segundo sentido, también decimos que objetos o procesos son transparentes cuando son visibles en todos sus detalles, abiertos a la inspección pública. Por ejemplo un proceso jurídico es transparente si toda la evidencia, todos los argumentos y todo el razonamiento de los jueces son accesibles al público. Gracias a la transparencia, la justicia no sólo se hace, sino que es evidente que se hace.

Si entendemos "transparencia de la mente" en el segundo de estos sentidos, podríamos hablar de una transparencia cartesiana. Cualquier cosa que pasa en la mente es transparente si el sujeto sabe (o por lo menos puede saber) que sucede. Descartes, al parecer, sostuvo que no es posible que suceda algo en mi mente sin que yo lo sepa.[5] Este tipo de transparencia *no* es el tipo del que hablan las teorías de la transparencia de la mente que recientemente han surgido para explicar nuestro autoconocimiento. Más bien, estas teorías usan el término de "transparencia" en el primero de los dos sentidos. La mente es transparente porque a través de ella se percibe el mundo y, percibiendo el mundo, también se puede llegar a conocer la mente. Para que sucediera esto a través de una observación adicional —observar la propia mente al percibir el mundo—, evidentemente sería necesario que la mente no fuera completamente transparente, porque no se puede observar lo que es perfectamente transparente y así camuflajeado de la observación.

Las teorías de la transparencia que a continuación se examinan se inspiran en una breve observación de Gareth Evans:

> Cuando hacemos una autoadscripción de creencia, nuestros ojos se dirigen, para decirlo así, o a veces literalmente, hacia afuera – al mundo. Si alguien me pregunta "¿Piensas que va a haber una tercera guerra mundial?", para responderle, tengo que atender precisamente a los mismos fenómenos exteriores que a los que atendería si estuviera contestando a la pregunta

[5] En una réplica a Arnauld, Descartes dice: "… en tanto que es una cosa pensante, no puede haber nada en la mente de lo que no esté consciente, esto me parece evidente" (Descartes, 1904, p. 246 [=AT.VII.246]; mi traducción). No es necesario interpretar esta cita como afirmando que el sujeto debe *saber* de cualquier suceso en su mente —entre otras cosas, eso depende de cómo Descartes entiende el término *conscia* (consciente)—, pero es una interpretación posible. Cfr. también Descartes (1904, p. 107 [=AT.VII.107]). Para una exégesis que *no* atribuye transparencia total a Descartes, cfr. Thomas (2009, pp. 26 y ss.).

> "¿Habrá una tercera guerra mundial?" Me pongo en una posición [apropiada] para contestar la pregunta de si creo que *p* poniendo en marcha el proceso (cualquiera que éste sea) mediante el cual respondo a la pregunta de si *p*. (Evans, 1982, p. 225).[6]

Evans aquí describe un caso donde se contesta una pregunta sobre la propia mente (¿qué es lo que crees?) a través de la respuesta a una pregunta sobre el mundo (¿habrá una guerra?). Este fenómeno ha llegado a ser nombrado "transparencia" —*for better or worse*, como dice Matthew Parrott (2015, p. e19)—. La idea detrás de esta terminología es que se puede conocer la propia mente, mientras que la atención se dirige al mundo. Y presumiblemente esto es así porque la mente es algo transparente que, sin embargo, figura en nuestra atención al mundo. Incluso, podría ser que la mente sea *completamente* transparente, es decir que sea *invisible* a cualquier observación, y sólo se *infiera* su existencia a partir de la observación del mundo.

En lo que sigue veremos cómo Alex Byrne y Jordi Fernández interpretan y desarrollan la observación de Evans para solucionar el problema del autoconocimiento, es decir, para explicar cómo es posible que nuestro autoconocimiento exhiba, al mismo tiempo, privilegio y peculiaridad.

4. Bel (Cree): la regla que Byrne sugiere para describir la adquisición del autoconocimiento

Alex Byrne sugiere que conocemos nuestras propias creencias porque seguimos cierta regla epistémica. Una regla epistémica describe cómo inferir conclusiones a partir de ciertos datos. El ejemplo general que Byrne propone es la regla Timbre:

> Timbre: Cuando suena el timbre, ¡cree que hay alguien en la puerta! (Byrne, 2005, p. 94).

La regla describe ciertas condiciones (que suena el timbre) y especifica algo que se puede creer bajo estas condiciones (que hay alguien en la puerta). Tiene la forma gramatical de un imperativo, porque recomienda creer algo (que hay alguien en la puerta) si las condiciones se cumplen.

[6] Todas las citas en este texto son traducciones mías del inglés, con la excepción de las citas de Descartes, que son traducciones del latín.

Según Byrne, seguimos una regla de este tipo si bajo las condiciones que la regla describe, creemos lo que la regla nos recomienda creer y – además – lo creemos *porque* se dan estas condiciones. Es decir, para seguir la regla, tenemos que darnos cuenta de que suena el timbre y con base en este hecho creer que hay alguien en la puerta (cfr. Byrne 2005: 94).

No parece demasiado difícil —se podría agregar a la sugerencia de Byrne— explicarse cómo podemos obtener una regla como Timbre para nuestro razonamiento. Lo principal es que la regla normalmente produce creencias verdaderas. Cuando suena el timbre, normalmente hay alguien en la puerta. Este hecho nos permite y conduce a formar una rutina en nuestro razonamiento que se puede describir como el seguimiento de la regla. También puede ser que alguien más nos enseña a seguir la regla y nos quedamos con ella porque produce resultados verdaderos. Dependiendo de nuestras inclinaciones externistas o internistas, podemos suponer más o menos entendimiento explícito de la finalidad de la regla. Este punto será de relevancia en la sección 5 más adelante, donde se evaluará una objeción de Matthew Boyle contra la teoría de Byrne.

Byrne sugiere que la observación de Evans sobre el autoconocimiento es correcta porque adquirimos tal conocimiento siguiendo una regla que es similar a Timbre, aunque, como veremos, tiene algunas propiedades especiales. La regla en cuestión se llama "Bel", derivado de *belief* en inglés, porque es concebida para describir la autoadscripción de creencias. Aquí la llamaré "Cree":

> Cree: Si *p*, ¡cree que crees que *p*! (Byrne, 2005, p. 95).

Similar a Timbre, la regla Cree describe ciertas condiciones (que es el caso que *p*) y especifica algo que se puede creer bajo estas condiciones (que yo creo que *p*). Nos recomienda creer que creemos que *p* si las condiciones se cumplen. Seguimos la regla si bajo las condiciones que describe creemos lo que según la regla podemos creer y lo creemos *porque* se dan estas condiciones. Esto significa que Cree requiere que nos demos cuenta (de la manera que sea) de que *p* y que, basado en este reconocimiento, formemos la creencia de que creemos que *p*.

Vale notar que Byrne también describe su regla como un esquema inferencial para la autoadscripción de creencias, el llamado "esquema doxástico":[7]

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

(Byrne, 2011b, p. 204).

En analogía con lo dicho anteriormente sobre la regla Cree, podemos decir que la premisa del esquema inferencial describe ciertas condiciones (que *p*) y la conclusión especifica algo que se puede creer bajo estas condiciones (que yo creo que *p*). Sólo se razona según el esquema doxástico si se reconoce que la premisa es verdadera (es decir, si uno se da cuenta de que *p*) y si con base en este reconocimiento se forma la creencia "Creo que *p*",[8] una creencia de segundo orden.[9]

Recordemos el ejemplo de Evans. Alguien nos pregunta si pensamos que va a haber una tercera guerra mundial. Según Evans, para contestar tenemos que atender a los mismos fenómenos exteriores a los que atenderíamos si estuviéramos contestando a la pregunta de si habrá una

[7] Como Byrne reconoce, esta noción es originalmente de André Gallois (1996, p. 46).

[8] Aquí, como en otras partes del texto, identifico una creencia por su contenido tal como el sujeto podría expresarlo. Si el sujeto cree que hay una manzana en la mesa, entonces podría expresar su creencia así: "Hay una manzana en la mesa". Por eso, si el sujeto cree que hay una manzana en la mesa, también escribo que tiene la creencia "Hay una manzana en la mesa", y si cree que cree que hay una manzana en la mesa, escribo que tiene la creencia "Creo que hay una manzana en la mesa".

[9] A lo largo de este texto supongo que la creencia de que *p* es una creencia sobre el mundo, no sobre un estado mental del sujeto, y en este sentido se trata de lo que llamo una *creencia de primer orden*. En contraste, si el sujeto cree que cree que *p*, tiene una creencia sobre un estado mental propio, y en este sentido se trata de lo que llamo una *creencia de segundo orden*. En principio, la variable "*p*" también podría representar una proposición sobre un estado mental del sujeto, haciendo su creencia de que *p* una creencia de segundo orden y su creencia de que cree que *p* una creencia de tercer orden. Pero aquí no tomo esta complicación en consideración y siempre presupongo que *p* es una proposición sobre el mundo.

tercera guerra mundial. Supóngase que concluimos que sí habrá una tercera guerra mundial. ¿Cuál es el siguiente paso? ¿Cómo llegamos de aquí a la *autoadscripción* de una creencia? Según Byrne, el siguiente paso es una inferencia que se puede describir como un seguimiento de la regla Cree o como una aplicación del esquema doxástico:

> Evans no contesta la pregunta explícitamente, pero la respuesta natural es que el siguiente paso involucra una *inferencia del mundo hacia la mente*: infiero que creo que habrá una tercera guerra mundial de la sola premisa de que habrá una (Byrne, 2011b, p. 203).

Vemos que en la teoría de Byrne la mente —por lo menos en tanto que se trata de creencias— es transparente en el sentido de no observado, sino inferido. Al conocer el mundo podemos conocer nuestras creencias, pero no observándolas, sino infiriéndolas. Sin embargo, se trata de una inferencia especial, ya que se infiere directamente de un hecho sobre el mundo ("*p*") hacia nuestra creencia de que este hecho se da ("Creo que *p*"), y no, como tal vez sea el caso cuando se trata de las creencias de *otras* personas, a partir del comportamiento de una persona hacia el hecho de que esta persona tiene cierta creencia.

Supongamos que el conocimiento de las propias creencias se adquiere por medio de la regla Cree, tal como Byrne sugiere. ¿Cómo ayudaría este hecho a solucionar el problema del autoconocimiento descrito al inicio de este texto? La idea de Byrne es que el seguimiento de Cree constituye un acceso peculiar y privilegiado a la propia mente. Veamos primero el acceso peculiar. El acceso a las propias creencias a través de Cree no es inmediato, sino por medio de una inferencia (como indica el esquema doxástico). Pero la inferencia es muy simple y no involucra ninguna percepción. En este sentido el acceso por medio de Cree es por lo menos comparativamente directo. El acceso también es peculiar en el sentido de ser exclusivo para uno mismo. Compárese la siguiente regla:

> Cree-3: Si *p*, ¡cree que Fred cree que *p*! (Byrne, 2005, p. 96).

Es obvio que Cree-3 puede resultar en muchas adscripciones falsas de creencia. Si yo me doy cuenta de que *p*, no se sigue que Fred cree que *p*. Es posible que Fred no se dé cuenta de lo mismo —tal vez porque esté mirando en otra dirección. Puede, me permito agregar a la consideración

de Byrne, que Cree-3 no sea una regla completamente inútil. Si no sé nada sobre Fred, puede ser una suposición sensata que Fred cree más o menos lo mismo que yo. Pero está claro que la confiabilidad de Cree-3 es mucho menor que la de Cree. Esta última es peculiar en el sentido de constituir un método especialmente para que uno pueda conocer sus propias creencias, no las de los demás.

Veamos ahora el acceso privilegiado. Como señalé, un sujeto sigue Cree si se da cuenta de que la condición antecedente descrita en la regla se cumple; es decir, el sujeto tiene que darse cuenta de que (es verdad que) *p*. Pero si una persona se da cuenta de que *p*, entonces esta persona *cree* que *p*. Darse cuenta de que *p* es una manera de formarse la creencia de que *p* o de activar una creencia (ocurrente o disposicional)[10] de que *p*. Ahora bien, si cumplir con el antecedente de la regla Cree significa que uno cree que *p*, entonces la autoadscripción que resulta del seguimiento de Cree necesariamente es verdadera, porque esta autoadscripción dice que el sujeto cree que *p*. Si sigue correctamente la regla Cree, entonces el sujeto hace una autoadscripción verdadera. En este sentido, Byrne dice que Cree es una regla *autoverificativa* (cfr. Byrne, 2005, p. 96).

¿Es el carácter autoverificativo de la regla Cree suficiente para explicar que tenemos un acceso privilegiado a las propias creencias? La regla Timbre, por ejemplo, no es autoverificativa. No es en virtud de oír el timbre, de darse cuenta de que suena, que es verdad que hay alguien en la puerta. La regla es buena en el sentido de conducente a producir creencias verdaderas porque la causa del sonido del timbre normalmente es que alguien en la puerta lo toca. Pero en principio es posible que exista otra causa del sonido del timbre y que no haya nadie en la puerta: puede haber un mal funcionamiento del timbre. No existe un mal funcionamiento análogo en el caso de Cree. Si me doy cuenta de que *p*, esto en sí ya constituye la formación de la creencia de que *p*. Cree provee un acceso privilegiado a las propias creencias en tanto que es una regla que no puede fallar por circunstancias inusuales en los que se presentan cadenas causales diferentes de las que existen en condiciones normales (cfr. Byrne, 2005, pp. 97 y ss.).

[10] Creemos muchas cosas sin que las estemos considerando en todo momento. Si creo que hay pingüinos en la Antártida sin que este contenido esté ocupando mi mente en este momento, la creencia es sólo *disposicional*. En cambio, si en este momento estoy juzgando o de otra forma considerando este contenido, mi creencia es *ocurrente*.

Aunque el carácter autoverificativo produce cierto tipo de acceso privilegiado, me parece que no explica la poca probabilidad de error en el momento de formarse creencias de segundo orden. Esto se puede apreciar en la siguiente regla para la autoadscripción de *conocimientos*:

Saber: Si *p*, ¡cree que sabes que *p*! (Byrne, 2012a, p. 190).

Saber también es una regla autoverificativa. Para seguirla correctamente, tenemos que reconocer o darnos cuenta de que la condición antecedente de la regla se cumple; es decir, tenemos que darnos cuenta de que *p*. Pero darse cuenta de que *p* es llegar a *saber* que *p*. Si no es verdad que *p*, no podemos darnos cuenta de que *p*. Así, quien sigue la regla correctamente, necesariamente hace una autoadscripción verdadera de un conocimiento. Pero por supuesto no tenemos un acceso privilegiado —en el sentido descrito inicialmente— a los propios conocimientos. Tal vez tengamos tal acceso a un componente psicológico de nuestros conocimientos (las creencias correspondientes, por ejemplo). Pero en general no tenemos más certeza acerca del hecho de que *sabemos* que *p*, que acerca del hecho de que *p* mismo. El problema es que —dependiendo de qué hecho de que *p* estamos hablando— puede ser muy fácil equivocarse acerca del antecedente de las reglas Cree y Saber. Y si nos equivocamos con *p*, también nos equivocamos con la autoadscripción que resulta de Saber ("Sé que *p*").

Pero si nos equivocamos acerca de *p*, ¿también resulta errónea la correspondiente autoadscripción de *creencia* ("Creo que *p*")? La respuesta es no, porque incluso si tomo *p* erróneamente como un hecho, es decir, si sólo me *parece* que *p* aunque no es verdad que *p*, todavía es el caso que formo la creencia de que *p* y, en consecuencia, la autoadscripción que resulta de mi aplicación errónea de Cree es verdadera. Byrne sugiere que en este caso digamos que el sujeto sólo *trata* de seguir la regla Cree, pero fracasa porque no tiene éxito en reconocer un cumplimiento de la condición antecedente de la regla (cfr. Byrne, 2005, pp. 97 y ss.). Cree se distingue de Saber por el hecho de que, a diferencia de Saber, también produce autoadscripciones verdaderas si el sujeto sólo trata de seguir la regla, pero fracasa en el sentido descrito. En la terminología de Byrne, ambas, Cree y Saber, son reglas autoverificativas, pero sólo Cree es una regla *fuertemente* autoverificativa (cfr. Byrne, 2011b, p. 206).

Con la autoverificación fuerte tenemos elementos significativos para explicar el acceso privilegiado a las propias creencias: La autoverificación

implica que Cree no depende de la cooperación causal del mundo para producir adscripciones correctas de creencia. Y la autoverificación *fuerte* significa que la aplicación de Cree no depende de ningún conocimiento del mundo.

¿Es *imposible* que Cree falle?[11] Byrne dice que no (cfr. Byrne, 2005, pp. 97 y ss.) y me parece que está en lo correcto. La aplicación de Cree puede fallar en el siguiente sentido: el sujeto se da cuenta (verídicamente o no) de que *p* y luego procede a formar la creencia correspondiente de segundo orden. Pero, en lugar de formar la creencia "Creo que *p*", llega a tener una creencia con otro contenido tal como "Creo que *q*". Por alguna razón, el sujeto no pudo retener el contenido "*p*" de su creencia inicial y utilizarlo nuevamente en la autoadscripción resultante de la aplicación de Cree. En este caso, su seguimiento de la regla Cree falla. Podríamos decir que el sujeto perdió el rastro del contenido ("*p*") de su creencia inicial y el resultado (la creencia "Creo que *q*") de su intento fallido de seguir Cree posiblemente es falso. O tal vez existan otras consideraciones por las que incluso la creencia "Creo que *q*" necesariamente es verdadera.[12] Pero si es así, no es porque el sujeto siguió la regla Cree. Sin duda, si un sujeto falla de esta forma en el seguimiento de la regla, entonces sufre de una falla de *racionalidad*.[13]

---

[11] Este y el siguiente párrafo no exponen ideas de Byrne, sino exclusivamente mías (aunque espero que Byrne estaría de acuerdo con lo expuesto). Para una posición contraria con respecto a la falibilidad de la regla Cree, cfr. Jordi Fernández (2013, pp. 19 y ss.). Fernández habla sobre la teoría de Richard Moran (2001), pero su argumentación parece aplicable a la de Byrne también.

[12] Tom Stoneham, por ejemplo, sugiere que la creencia de que creo que *p* necesariamente es verdadera (y en consecuencia infalible) porque "contiene" la creencia de que *p* como elemento constitutivo (cfr. Stoneham, 1998). Stoneham no explica cómo llegamos a formar creencias de segundo orden. Es por lo menos concebible que las formamos siguiendo la regla Cree, pero que los resultados de esta formación de creencia siempre son verdaderas (incluso cuando fallamos en seguir la regla Cree) debido a relaciones constitutivas como las descritas por Stoneham.

[13] La tesis defendida aquí es que un sujeto racional que tiene la regla Cree a su disposición tiene un acceso privilegiado y peculiar a sus propias creencias. Es decir, tiene un método especial a su disposición para adquirir conocimiento de sus creencias. Sydney Shoemaker defiende una tesis más fuerte: que es una consecuencia de nuestra racionalidad que no podemos tener (por lo menos

¿Qué tan probable es que fallemos de esta manera en seguir la regla Cree? La inferencia —si por el momento interpretamos Cree según el "esquema doxástico"— de "*p*" a "Creo que *p*" es extremadamente simple. La capacidad de retener un contenido y utilizarlo nuevamente es esencial para poder hacer inferencias de cualquier tipo. Por ejemplo, inferir "*p*" de "*p y q*" requiere retener el contenido "*p*". Nuestro acceso a las propias creencias debería ser tan privilegiado como nuestra capacidad para hacer inferencias simples sin perder el rastro y corromper los contenidos de las inferencias.[14] Ciertamente somos mejores al retener y reutilizar contenidos en inferencias simples que al hacer inferencias más complejas, porque hacer inferencias complejas requiere retener contenidos y *además* lidiar con la complejidad de la inferencia, es decir sacar las conclusiones correctas a pesar de la multitud de premisas, la variedad de sus relaciones justificativas con la conclusión, etc. ¿Somos también mejores en retener y reutilizar contenidos en inferencias simples que en conocer el mundo a través de la percepción o el testimonio? Me parece que se puede decir que, por lo menos en general, sí. Es muy fácil pensar en fallas de percepción: por ejemplo, Descartes menciona que a menudo la percepción falla cuando los objetos a percibir son "muy pequeños o distantes" (Descartes, 1904: 18 [=AT.VII.18]). Similarmente puede suceder que un testimonio no sea confiable: alguien me cuenta una mentira y le creo. Pero veamos una inferencia simple. Si Pedro me dice que va al restaurante chino o al de mariscos, y más tarde no lo encuentro en el de mariscos, infiero que está en el restaurante chino. Esta conclusión puede fallar de muchas maneras —tal vez Pedro tuvo un accidente, por ejemplo—; pero parece poco probable que falle por una falta mía en retener y utilizar los contenidos relevantes, tales como "estar en el restaurante chino" y "estar en el restaurante de mariscos". Una falla de este tipo sería si de mis premisas ("Pedro está en el restaurante chino o en el de mariscos" y "Pedro no está en el restaurante de mariscos") concluyo que Pedro está en el restaurante italiano porque no logro retener el contenido "estar en el restaurante chino" y el contenido se me corrompe para formar uno nuevo: "estar en el restaurante italiano". Me

ciertas) creencias sin saber que las tenemos, porque tener el autoconocimiento es constitutivo (en criaturas racionales como nosotros) de tener las creencias de primer orden (cfr. Shoemaker, 1994, pp. 288 y ss.). He discutido algunos de los argumentos de Shoemaker en Fricke (2012).

14 Cfr. Fricke (2009).

parece que tales fallas, aunque no son imposibles, no son muy comunes. Tal vez haya casos simples donde la percepción puede alcanzar el mismo grado de confiabilidad que nuestra capacidad para retener y reutilizar contenidos. Pero en general, la percepción y, especialmente, el testimonio es mucho más susceptible a fallar en la producción de conocimientos. Este hecho junto con la autoverificación fuerte es la razón por la que autoadscribirse creencias a través de la regla Cree parece explicar el acceso privilegiado que tenemos a nuestras creencias.

Hemos visto cómo la regla Cree de Byrne implica tanto un acceso peculiar como un acceso privilegiado a las propias creencias. En lo que sigue, discutiré dos objeciones a la teoría de Byrne.[15]

5. Dos objeciones a la teoría de Byrne

Las dos objeciones a la teoría de Byrne que a continuación presentaré pueden distinguirse según si critican el aspecto normativo o metafísico del seguimiento de la regla Cree. La primera objeción es de Matthew Boyle; la segunda, de Brie Gertler. Expondré las dos objeciones y sugeriré algunas respuestas que —me parece— Byrne podría hacer para defender su propuesta. Las respuestas son mías, pero espero que sean en el espíritu de la teoría de Byrne.

Matthew Boyle dice que la inferencia (implícita en Cree) por medio de la cual autoadscribimos nuestras creencias según Byrne es "absurda" *(mad)*.[16] En general, la inferencia de "*p*" a "Creo que *p*" ni es deductivamente válida ni inductivamente fuerte (cfr. Byrne, 2011b, p. 204). La mayoría de los hechos en el mundo existen sin que yo lo crea. No es el caso, entonces, que el hecho de que *p* muestre de alguna manera el que yo creo que *p*. En consecuencia, dice Boyle, si me doy

[15] Se trata probablemente de las objeciones más prominentes que han surgido. Una crítica que no discutiré aquí aunque sí merece ser examinada es la de Markos Valaris (2011), que reclama que la supuesta regla epistémica Cree no es compatible con el razonamiento hipotético donde se supone que *p* para luego deducir una contradicción y así refutar que *p*. (Si supongo que *p*, Cree debería permitirme concluir que creo que *p*. Pero si sucede que de hecho no creo que *p*, por este mismo hecho ya quedaría refutada la suposición de que *p*.) La pregunta es, entonces, qué tipo de regla es Cree, considerando que no puede ser usada en el razonamiento hipotético.

[16] *Mad* también puede ser traducido como "loco", "demente", "disparatado" o "sin sentido" (entre otras posibilidades).

cuenta de que mi única razón para juzgar que *creo* que *p* es el hecho de que *p*, debería abandonar la creencia de segundo orden porque le falta una justificación adecuada. "[U]n mínimo de entendimiento racional me informará que, incluso si es verdad que *p*, eso en sí no tiende a mostrar que yo lo creo" (Boyle, 2011, p. 230).

Boyle no cuestiona que la regla Cree generalmente produce autoadscripciones verdaderas de creencia. Pero estipula una condición fuertemente internista sobre la manera en que podemos hacerlas: debo basarme en algo que *muestra* la verdad de la autoadscripción. La teoría de Boyle evidencia que no es fácil encontrar tal base para nuestras autoadscripciones. Su solución es la idea de que las creencias de primer orden desde su inicio van junto con nuestro conocimiento de que las tenemos. Normalmente este autoconocimiento es tácito. Pero si reflexionamos sobre nuestras creencias, puede llegar a ser explícito en una autoadscripción. No me detengo en la propuesta positiva de Boyle.[17] Pero ¿qué podría Byrne responder a las críticas de Boyle?

[17] La crítica de Boyle que discuto aquí es que la inferencia que según Byrne nos proporciona el autoconocimiento no tiene sentido. Sin embargo, su propuesta positiva (la cual no discuto) también constituye una crítica a Byrne. Es una consecuencia de esta propuesta que no es posible que una proposición sobre el mundo como "*p*" sea reconocida como verdadera y que luego todavía le quede al sujeto una tarea *adicional* por realizar, a saber: la de formar una creencia de segundo orden de que cree que *p*. Según Boyle, reconocer que *p* o darse cuenta de que *p* siempre conlleva, por lo menos tácitamente, un conocimiento de segundo orden, como "Sé que *p*" o "Creo que *p*". La creencia de primer orden y el conocimiento de ella son aspectos de un solo estado, según este autor (cfr. Boyle, 2011, p. 228). Un árbitro me ha señalado que esta afirmación de Boyle se puede respaldar fenomenológicamente, que el hecho de que *p* siempre se muestra al sujeto como parte de una perspectiva en primera persona. Si esto es así, el punto de partida del procedimiento de Byrne, una creencia de que *p* sin autoconocimiento todavía, no existe y por lo tanto el procedimiento no puede explicar nada. Para evaluar esta objeción habría que examinar en qué, exactamente, la perspectiva de la primera persona o el conocimiento tácito de segundo orden de Boyle consiste y, en especial, si en realidad constituye un autoconocimiento. Un filósofo como Byrne probablemente contestaría que es plausible que ciertos animales tienen creencias de primer orden (creencias de que *p*), pero no creencias de segundo orden (creencias de que creen que *p*) y que esto muestra que no es imposible estar en la posición de creer que *p* y usar esta premisa para inferir con la ayuda de Cree una creencia adicional de que uno cree que *p* (sobre esta respuesta, cfr. Boyle, 2011, p. 228, nota al pie 5).

Una respuesta podría señalar que la condición internista de Boyle es demasiado fuerte. No necesariamente descartamos una creencia solamente porque no encontramos nada que *muestra* la verdad de la creencia. Al contrario, tal vez sea más razonable sostener las propias creencias por lo menos hasta encontrar razones en contra.

Otra respuesta posible es que no es tan difícil que la condición internista (o algo semejante) se cumpla. Se podría argumentar que el sujeto sí puede tener algún tipo de entendimiento de por qué Cree produce autoadscripciones verdaderas de creencia. Si algo es verdadero (a mi juicio), entonces formo la creencia de que lo es. En tanto que soy racional, si me parece que algo es verdadero, entonces formo una creencia de ello.[18] En consecuencia, una inferencia de Byrne a partir de lo que es verdad (cuando yo considero la pregunta) necesariamente revela qué es lo que creo. La necesidad de esta afirmación reside en mi racionalidad como sujeto epistémico.

La segunda objeción contra la teoría de Byrne es que su método para llegar a autoadscripciones de creencia en realidad no *revela* creencias de primer orden, sino que *genera* nuevas creencias. Las autoadscripciones que hacemos a través de Cree son verdaderas, pero no lo son porque ya teníamos las creencias adscritas desde antes.

Esta objeción ha sido desarrollada en detalle por Brie Gertler (2011). Ella señala, por ejemplo, que podría ser que no tengo ninguna creencia acerca de si *p*. Si ahora uso Cree para averiguar si creo que *p* o no, entonces me pregunto si es verdad que *p*. Para responder considero la evidencia a favor y en contra de *p* y es posible que forme alguna creencia acerca de si *p*. Supóngase que formo la creencia de que *p*. Siguiendo Cree, en seguida formo la creencia de segundo orden: "Creo que *p*". Por las razones que he explicado anteriormente, esta autoadscripción de creencia va a ser verdadera. Pero lo es solamente porque la creencia de primer orden ("*p*") se formó en respuesta a la aplicación de Cree. La

[18] Richard Moran caracteriza esta afirmación "como una Suposición Trascendental del Pensamiento Racional" (*something like a Transcendental assumption of Rational Thought*), y la formula así: "lo que en realidad creo sobre X se puede determinar, hacer verdad, a través de mi reflexión sobre X mismo" (Moran, 2003, p. 406). Según Moran, esta Suposición Trascendental es lo que me da el derecho epistémico de seguir el procedimiento de Evans. Hay que mencionar, sin embargo, que, a diferencia de Byrne, Moran no interpreta el procedimiento de Evans como una inferencia.

regla no detectó la ausencia inicial de la creencia de que *p*. Es claro que Cree no puede ser usado para autoadscribirse la falta de una creencia: la condición para la cual Cree hace una recomendación es que el sujeto se dé cuenta de algún hecho de que *p*. Pero si el sujeto no tiene opinión alguna sobre si *p* o no, entonces no se cumple esta condición para la ejecución de Cree.

Gertler razona que un problema similar ocurre con creencias actuales (*occurrent beliefs*). Supóngase que trato de utilizar Cree para averiguar si actualmente creo que *p*. Primero me pregunto si *p*, me doy cuenta de que sí es verdad que *p* y luego correctamente autoadscribo esta creencia ("Creo que *p*"). El problema que Gertler observa es que cuando me pregunto si *p*, como primer paso de la aplicación de Cree, esta pregunta podría ser lo que me causa, tal vez después de una nueva revisión de la evidencia disponible, a formar la creencia de que *p*. Eso podría ser así incluso si antes de aplicar Cree yo no creía que *p* o creía que no *p*. En este caso, aunque la autoadscripción resultante es verdadera, no revela o detecta una creencia actual que ya existía antes de la aplicación de Cree. Más bien, la aplicación genera una nueva creencia y luego la autoadscribe.

Supóngase ahora que queremos saber si tenemos una creencia *disposicional* de que *p*, una creencia que no figura en nuestros pensamientos actuales, pero que puede ser activada inmediatamente si surge la pregunta correspondiente. Gertler señala que si queremos averiguar si tenemos esta creencia utilizando Cree es importante tener cuidado de no cambiar la creencia disposicional al introducir nueva evidencia. Aplicando Cree, nos preguntamos si es verdad que *p*. Si contestamos esta pregunta considerando la evidencia que tenemos *ahora*, es posible que lleguemos a una respuesta distinta de lo que ya creemos disposicionalmente porque la creencia disposicional se formó anteriormente cuando tal vez no toda la misma evidencia que tenemos ahora estaba disponible. La aplicación de Cree sólo revelará la creencia disposicional si nos cuidamos de no tomar en consideración la nueva evidencia. Si de esta manera sólo nos basamos en lo que ya tenemos en mente, lo "interno", nuestra respuesta a la pregunta de si *p* probablemente sí reflejará nuestras creencias disposicionales. Pero esto significa que en este caso la aplicación de Cree *presupone* algún tipo de autoconocimiento, a saber, la distinción entre lo que "se cree internamente" y lo que es verdad en vista de influencias "externas".

¿Qué tan seria es la objeción de Gertler? Lo primero que quisiera hacer notar es que se podría aceptar la objeción sin que esto afectara el núcleo de la teoría de Byrne. La idea más importante de esta teoría es que el seguimiento de la regla Cree es una manera plausible de adquirir con privilegio y peculiaridad autoconocimiento sobre las propias creencias. Y esta idea no se cuestiona por el argumento. Sólo se cuestiona si el método detecta creencias actuales o disposicionales que ya tenemos antes de ejecutar el método, y tal vez sea apropiado ser escéptico sobre la idea de que podemos tener autoconocimiento privilegiado y peculiar de estas creencias anteriores. Pero, segundo, no es necesario interpretar la aplicación de la regla Cree tal como lo sugiere Gertler. En particular, no es necesario para tal aplicación hacerse explícitamente preguntas tales como "¿Es verdad que *p*?". Más plausiblemente, hacer uso de la regla podría consistir simplemente en una tendencia a inferir a partir de los hechos con los que nos encontramos que las creemos. En este caso no necesariamente habría preguntas que nos estimulan a evaluar nuevamente nuestra evidencia. La regla simplemente se sigue porque está a nuestra disposición como una tendencia de hacer inferencias a partir de lo que se presenta como verdadero. Aquí el resultado sí consiste en la autoadscripción de creencias que ya tuvimos antes de poner la regla Cree en práctica.

6. Bypass: el procedimiento que Fernández sugiere para describir la adquisición del autoconocimiento

Jordi Fernández (2013) propone una teoría interesante del autoconocimiento que desarrolla la idea de la transparencia de una manera un poco diferente que la teoría de Byrne. En lo que sigue examinaré la propuesta de Fernández para poder evaluar si es preferible a la de Byrne. La principal diferencia entre las dos teorías es que, según Fernández, las autoadscripciones de creencia se basan en los mismos estados mentales que también son el fundamento (*grounds*)[19] de las

[19] Fernández usa el término *grounds for belief* para referirse a aquellos estados mentales que tienden a causar la creencia en cuestión. Tal vez la traducción más indicada de *grounds* sería "causa" porque, según Fernández, el término no tiene connotaciones normativas y sólo se caracteriza por su rol causal (cfr. Fernández, 2013, p. 45, nota de pie 7). Sin embargo, en un resumen de su libro, escrito en español, el autor mismo traduce (o por lo menos autoriza la

creencias de primer orden que aquí se autoadscriben. En cambio, como vimos, según Byrne, las autoadscripciones se basan en las creencias de primer orden mismas, no en el fundamento de estas creencias. Veamos.

Una idea central de la teoría de Fernández es que normalmente formamos nuestras creencias con base en algún otro estado mental y existe una regularidad entre el tipo de base o fundamento y el tipo de creencia que formamos. Por ejemplo, si el estado mental es la percepción de una manzana en frente de mí, entonces normalmente formo la creencia de que hay una manzana en frente de mí. Ahora, la idea ingeniosa de Fernández es la siguiente: si el mismo tipo de estado mental regularmente produce —"fundamenta", diría Fernández— el mismo tipo de creencia de primer orden, entonces también podemos basarnos en este tipo de estado para *autoadscribirnos* las creencias que el estado normalmente produce. Si un estado *E* normalmente causa una creencia *C* de primer orden, entonces cuando tenemos el estado *E* también podemos formar la creencia de que tenemos *C* (una creencia de segundo orden) con base en *E*. Porque siempre (o por lo menos siempre en circunstancias normales) que tenemos *E* también se produce la creencia *C*, podemos utilizar *E* como indicador de la creencia *C*. Es decir, podemos autoadscribir *C* en una creencia de segundo orden que se basa en *E*. En el ejemplo anterior, cuando me parece que veo una manzana en frente de mí, normalmente formo la creencia (de primer orden) de que hay una manzana en frente de mí. Por consiguiente, sugiere Fernández, cuando me parece que veo una manzana en frente de mí, también puedo formar la creencia de segundo orden "Creo que hay una manzana en frente de mí". Esta creencia de segundo orden no está basada directamente en la creencia de primer orden ("Hay una manzana en frente de mí"). Más bien, se "sobrepasa" (*bypass*) la creencia de primer orden y la creencia de segundo orden se basa en la percepción que también es el fundamento de la creencia de primer orden.[20]

Fernández trata de capturar la regularidad entre nuestras creencias y sus fundamentos en las siguientes afirmaciones generales:

---

traducción de) *grounds* como "fundamentos", y añade: "Los fundamentos que un sujeto tiene para una creencia se estipula [en mi libro] que son estados mentales que tienden a causar que ese sujeto tenga la creencia en cuestión" (Fernández, 2015a, p. 96). En lo que sigue, usaré "fundamento" donde Fernández usa *grounds*.

[20] Este párrafo expone ideas de la sección 2.4. de Fernández (2013, pp. 48 y ss.).

Para cualesquiera proposiciones P, Q y cualesquiera sujetos S, S*:

> (i) Si S aparentemente[21] percibe que P, entonces S llega a creer que P.
>
> (ii) Si S aparentemente recuerda que P, entonces S llega a creer que P.
>
> (iii) Si S cree que S* le está dando la información de que P, entonces S llega a creer que P.
>
> (iv) Si S cree que Q y S cree que P se sigue de Q, entonces S llega a creer que P. (Fernández, 2013, p. 46).
>
> (v) Si a S le parece intelectualmente que P, entonces S llega a creer que P. (Fernández, 2013, p. 48).

En cada una de estas afirmaciones condicionales, se especifica un estado mental en el antecedente, el cual normalmente constituye el fundamento para la formación de una creencia, la cual se especifica en el consecuente. Las afirmaciones (i) a (iv) describen los estados mentales asociados a la percepción, la memoria, el testimonio y el razonamiento. La afirmación (v) describe un estado mental que constituye el fundamento para una creencia a priori. En cada caso también se describe la creencia que normalmente se forma con base en estos estados mentales antecedentes. No me detendré en los detalles de las cinco afirmaciones. El punto importante es que se supone que existe una regularidad en nuestras maneras de formar creencias y ésta puede ser descrita en enunciados como los de (i) a (v).

[21] ¿Por qué Fernández habla de una percepción *aparente*? La razón (no muy elaborada por Fernández) es que el sujeto forma su creencia de que *p* en respuesta a una experiencia que le presenta el mundo como si *p*. Esta experiencia puede ser verídica cuando el sujeto percibe que *p*, o puede ser falsa cuando se trata sólo de una percepción *aparente*. Lo importante para Fernández es que un mismo tipo de estado mental (la experiencia que presenta el mundo como si *p*) tiende a causar en el sujeto la creencia de que *p*. Consideraciones análogas aplican a la afirmación (ii).

El procedimiento de Bypass se justifica gracias a esta regularidad. Fernández describe Bypass así:

> Para cualquier proposición P y sujeto S: Normalmente, si S cree que cree que P, entonces hay un estado E tal que
>
> (a) la creencia de orden superior ha sido formada con base en E.
>
> (b) E constituye el fundamento de la creencia de que P, de S. (Fernández, 2013, p. 49).

Bypass es la interpretación que Fernández ofrece de la observación de Evans. Si me preguntan si creo que habrá una tercera guerra mundial, examino mi evidencia para tal guerra. Si hay percepciones, recuerdos, testimonios, etc., que respaldan el que habrá una tercera guerra mundial, autoadscribo con base en estos estados mentales la creencia sobre la guerra: "Creo que habrá una tercera guerra mundial". Mi autoadscripción no se basa en la creencia de primer orden, sino en la evidencia (o las razones) para tal creencia de primer orden. La creencia misma se sobrepasa, tal como dice el nombre del procedimiento: Bypass.

La teoría de Fernández tiene un aspecto metafísico y uno normativo.[22] El aspecto metafísico describe la mecánica causal de la formación de creencias; el aspecto normativo, las relaciones de justificación. Fernández se concentra en las relaciones normativas, pero para mi discusión de algunas objeciones en la siguiente sección conviene distinguir los dos aspectos. La metafísica de Bypass consiste en varias relaciones causales entre los fundamentos, las creencias de primer orden y las creencias de segundo orden. Los fundamentos son estados mentales asociados con percepción, memoria, testimonio y razonamientos. Estos estados mentales regularmente causan la formación de creencias de primer orden y, según Bypass, también de segundo orden. Las creencias de segundo orden son causadas por los mismos fundamentos que las creencias de primer orden que se autoadscriben en aquellas. Así, la percepción de

[22] La distinción de estos dos aspectos, aplicada a la teoría de Fernández, es mía. Este autor no la usa como tal.

una manzana en frente de mí causa tanto la creencia de primer orden "Hay una manzana en frente de mí", como la creencia de segundo orden "Creo que hay una manzana en frente de mí".

Veamos ahora el aspecto normativo de la teoría. El estado mental que constituye el fundamento de una creencia de primer orden justifica esta creencia en virtud de su contenido representacional. La percepción de una manzana en frente de mí justifica mi creencia de que hay una manzana en frente de mí porque tiene el mismo contenido representacional. La percepción de una manzana en frente de mí también justifica mi creencia de que *creo que* hay una manzana en frente de mí. Pero no la justifica en virtud de su contenido representacional, sino en virtud de la relación causal entre la percepción y la creencia de primer orden. Existe una regularidad entre las dos: la percepción regularmente causa la creencia. Es por la existencia de tal regularidad que está justificada la autoadscripción de la creencia de primer orden con base en la percepción (cfr. Fernández, 2013, pp. 53 y ss. [sección 2.5]).

Nótese que es posible que la creencia de primer orden no tenga justificación aunque la autoadscripción de ésta sí. Podría ser, por ejemplo, que regularmente razono falazmente afirmando el consecuente.[23] En este caso, mi creencia de primer orden carecería de justificación.[24] Sin embargo, si autoadscribiera la creencia resultante de mi razonamiento falaz con base en los estados mentales que me hacen embarcar regularmente en la afirmación del consecuente, entonces haría una autoadscripción no sólo verdadera, sino también justificada. La razón es, según Fernández, que incluso en este caso existe una regularidad causal entre un estado mental fundamento y la creencia de primer orden. Se trata de una regularidad irracional que no sirve para justificar la creencia de primer orden. Pero esta irracionalidad no es

[23] El ejemplo es mío. Fernández discute un caso análogo (la creencia de primer orden no tiene justificación, la creencia correspondiente de segundo orden sí), pero no usando falacias lógicas, sino prejuicios, como ejemplo de creencias de primer orden sin justificación (cfr. Fernández, 2013, pp 65 y ss.).

[24] Estoy suponiendo, esquemáticamente, que mi razonamiento se basa en las creencias de que q y de que si p, entonces q. Con base en estas creencias concluyo, directa y falazmente, que p. Este esquema no coincide con el principio (iii) citado anteriormente con el que Fernández trata de describir nuestro razonamiento inferencial. Para más discusión véase, más adelante, la tercera observación a la segunda objeción en contra de la teoría de Fernández.

relevante cuando la cuestión es si la autoadscripción tiene justificación. La tiene simplemente en virtud de la regularidad entre el estado mental y la creencia (irracional) de primer orden.

¿Cómo soluciona la teoría de Fernández el problema del autoconocimiento? Es decir, ¿cómo explica su teoría que nuestro autoconocimiento exhibe tanto un acceso peculiar como uno privilegiado a nuestras creencias de primer orden (cfr. sección 2 *supra*)? Veamos primero el acceso peculiar.[25] Claramente Bypass es un procedimiento sólo para la adscripción de creencias a uno mismo. Los estados mentales que son fundamento de mis creencias de primer orden no justifican la adscripción de creencias a otras personas porque no existe una regularidad causal entre *mis* estados mentales, por ejemplo mis percepciones, y las creencias de *otras* personas. Es posible que una percepción mía haga un tanto más probable que otra persona —especialmente si está cerca y en una posición similar a mí— también tenga las creencias que se fundamentan en mi percepción. Pero la probabilidad de que tal conclusión sea equivocada también es mucho más grande que en el caso de mis autoadscripciones. Por lo mismo, Bypass es principalmente un método para conocer las propias creencias, no las de otras personas. Bypass también explica por qué el acceso a las propias creencias es más directo que el acceso a las creencias de otras personas. Bypass no es una inferencia, sino la formación de una creencia en respuesta inmediata al estado mental que lo justifica. No es necesario percibirse a sí mismo para ejecutar Bypass, pero muchas veces sí es necesario observar a otra persona para poder adscribirle creencias. En este sentido, Bypass es un procedimiento más directo para conocer (las propias) creencias.

¿Cómo explica Bypass el acceso privilegiado[26] que tenemos a las propias creencias? La idea general de Fernández es que la adscripción

[25] La noción de "acceso peculiar" que utilizo es similar a la noción de "acceso especial" de Fernández. Él dice que el acceso especial consiste en el hecho de que el sujeto no depende de razonamientos ni de evidencia de comportamiento en sus autoadscripciones de actitudes proposicionales (cfr. Fernández, 2013, p. 5).

[26] La noción de acceso privilegiado que se usa aquí es muy similar a la noción de "acceso fuerte" de Fernández (cfr. 2013, p. 6). Fernández también usa el término "acceso privilegiado", pero en sus textos la expresión comprende tanto lo que él llama "acceso especial" como "acceso fuerte".

de creencias a otras personas "depende de la confiabilidad de algunas facultades de la cual no dependen mis autoatribuciones [de creencia]" (Fernández, 2013, p. 59). Para saber qué creen otras personas tengo que observar su comportamiento, incluso su comportamiento verbal, y sacar mis conclusiones a partir de estas observaciones. Si eso es correcto, entonces el acceso a las creencias de otras personas depende de facultades como la percepción y de mi capacidad para hacer inferencias a partir de la percepción. En consecuencia, este acceso está sujeto a posibles fallas que puedan ocurrir en la percepción o en mis inferencias. El procedimiento de Bypass, en cambio, no depende de la percepción verídica y tampoco requiere hacer inferencias. Sí, se requiere que *tenga* el estado mental que es el fundamento para mi creencia de primer orden, a saber, una experiencia perceptiva, un (aparente) recuerdo, la aceptación de un testimonio, etc. Pero no importa si se percibe o recuerda correctamente o si el testimonio es verdadero. La autoadscripción de creencia, según Bypass, se basa directamente en ese estado mental. No se necesita hacer una inferencia. Por consiguiente, el procedimiento de Bypass no está sujeto a fallas en la percepción, memoria, testimonio o en mis inferencias. Se podría preguntar aquí qué tan confiable es la formación de una creencia de segundo orden con base en un estado mental como una percepción, recuerdo o testimonio. Fernández no considera esta interrogante y no es fácil decir cómo se podría contestar. Pero su explicación del acceso privilegiado depende de que esta formación de creencias de segundo orden sea un ejercicio más fácil que las percepciones e inferencias necesarias para adscribir creencias a otras personas.

7. Dos objeciones a la teoría de Fernández

Me parece que las objeciones de Boyle y Gertler contra Byrne, que discutí en sección 5, pueden ser reformuladas de tal forma que también apliquen a la teoría de Fernández. Sin embargo, no las examinaré aquí, porque creo que también las respuestas que Fernández podría dar a estas objeciones son similares a las que podría dar Byrne y tienen una plausibilidad similar.

En lugar de las ideas de Boyle y Gertler, quiero desarrollar dos objeciones nuevas que todavía no han sido investigadas en la discusión de la teoría de Fernández. Ambas tienen que ver con la pregunta de si su teoría puede explicar el conocimiento de las propias creencias cuando

nuestra manera de formarlas cambia. La primera objeción, que concierne el aspecto normativo de la teoría, dice que el procedimiento de Bypass no es capaz de justificar autoadscripciones de creencias que se formaron accidentalmente. La segunda, que concierne el aspecto metafísico, pone en duda si el procedimiento de Bypass puede responder adecuadamente a cambios algunos en nuestras maneras de formar creencias.

La primera objeción surge del hecho de que muchas veces nuestras creencias de primer orden son irracionales. La objeción señala que no es plausible que en algunos casos el autoconocimiento acerca de tales creencias carece de justificación, como parece implicar la teoría de Fernández. Ya vimos anteriormente un ejemplo de la formación irracional de creencias: supóngase que regularmente afirmo el consecuente. En este caso, formo creencias de primer orden de manera irracional. Sin embargo, Fernández argumenta, la *autoadscripción* de tales creencias a través del procedimiento de Bypass *no* es irracional, sino justificada porque existe una *regularidad* en mi manera de formar las creencias de primer orden. La regularidad es irracional; regularmente razono falazmente y afirmo el consecuente. Pero aun así, la regularidad en razonar de esta manera falaz justifica el procedimiento de Bypass. Es porque *siempre* formo mis creencias de primer orden sobre estos fundamentos irracionales que también puedo autoadscribirlas sobre estos fundamentos.

Me parece que el argumento de Fernández es correcto. Pero el problema es que a veces formamos creencias de una manera irracional e *irregular*. Puede ser que normalmente no afirmo el consecuente, ya que sé que es un razonamiento falaz. Pero en una ocasión aislada, tal vez por estar distraído por alguna razón, sí cometo la falacia. La creencia que formo entonces es el resultado de un error no habitual, sino accidental e inusual. ¿Puedo saber de la creencia que formo de esta manera? No parece plausible negar que una creencia propia, que es accidentalmente irracional, se pueda conocer en exactamente la misma forma que una creencia racional. Pero esto es lo que dice la teoría de Fernández. Tal vez sí pueda utilizar el procedimiento de Bypass para hacer una autoadscripción de la creencia de primer orden (aunque véase la segunda objeción más adelante). Pero está claro que Bypass no puede justificarse, en este caso, por la regularidad con la que el sujeto siempre forma este tipo de creencias de primer orden. Esa regularidad no existe en nuestro caso porque el error es atípico e inusual para mí, no un hábito regular. El uso de Bypass para la autoadscripción de una creencia que

formo de manera atípicamente irracional no se puede justificar de la misma manera como en otros casos donde la creencia de primer orden se forma según alguna regularidad. En consecuencia, Bypass no puede darnos conocimiento de las propias creencias si las formamos a través de un proceso accidentalmente erróneo.

Nótese que para el anterior argumento no es relevante que la creencia de primer orden sea errónea.[27] Lo que impide que la autoadscripción tenga justificación es únicamente el hecho de que la creencia de primer orden se formó *accidentalmente*, es decir sin regularidad. Se puede formular un argumento análogo al del párrafo anterior con una creencia de primer orden *verdadera*, pero formada accidentalmente. Podríamos pensar, por ejemplo, en el sujeto que tiene el hábito de afirmar falazmente el consecuente. Pero en una ocasión el sujeto se equivoca y accidentalmente razona correctamente, digamos siguiendo el esquema de *modus ponens* en lugar de afirmar el consecuente. El problema para Fernández es que en este caso no existe una regularidad con la que el estado mental (aquí: la creencia de que *p* y la creencia de que *q* se sigue de *p*) causa la creencia de primer orden en el sujeto. En consecuencia, seguir el método de Bypass y formar la creencia de segundo orden con base en el estado mental no tiene justificación. La objeción contra Fernández es que esta consecuencia de su teoría no es plausible. ¿Por qué no debería ser posible conocer una creencia propia, independientemente de si la formé accidentalmente o no?

En lo que sigue, me concentraré en el caso donde la creencia de primer orden formada accidentalmente es errónea. Pero los argumentos se pueden reconstruir análogamente para los casos donde la creencia accidental es verídica.

Fernández puede contestar mi crítica señalando su definición de lo que es el "fundamento de una creencia":

> Sea 'S' un sujeto, 'G' un estado [mental] y 'B' una creencia. Utilizaré las expresiones 'G constituye el fundamento [*grounds*] de B en S' y 'B está fundamentado [*grounded*] sobre G en S' para referirme al hecho de que S tiende a tener B cuando está en G (Fernández, 2013, p. 45).

[27] Agradezco a un árbitro de *Tópicos* por haber llamado mi atención a este punto.

Una consecuencia de esta definición de "fundamento" es que no puede haber fundamentos irregulares de nuestras creencias. Un estado mental sólo puede ser el fundamento para una creencia si *tiende* a producir la creencia en el sujeto. Si un sujeto tiende a razonar falazmente existe un estado mental en el sujeto que tiende a producir las creencias falaces. Pero si el sujeto cae en un error aislado que no corresponde a algún hábito regular, evidentemente los estados que producen el error no *tienden* a producir la creencia errónea en el sujeto. Más bien, parece que, según la definición de "fundamento" que ofrece Fernández, creencias que son atípicamente erróneas o solamente se formaron accidentalmente *no tienen fundamento alguno* en el sujeto. Y si no tienen fundamentos, el procedimiento de Bypass no nos puede dar conocimiento de ellas, ya que Bypass describe autoadscripciones de creencia que se basan en los fundamentos de las creencias de primer orden. Si Bypass no tiene aplicación en la autoadscripción de estas creencias accidentales, tampoco puede faltarle una justificación al procedimiento en este caso.

En este punto quiero hacer tres observaciones críticas. Las primeras dos son más específicas y sobre la pregunta de si podemos conocer las propias creencias cuando estás son accidentalmente erróneas. La tercera es una crítica más general a la teoría de Fernández.

La primera observación crítica es que la definición que Fernández ofrece de "fundamento" puede parecer arbitraria en el punto que aquí es relevante. Según el autor, la noción no "conlleva ninguna connotación normativa. Los fundamentos se conciben aquí como estados que se caracterizan simplemente por sus roles causales" (Fernández, 2013, p. 45, nota de pie 7). El fundamento de una creencia es, entonces, aquel estado mental que tiende a causar la creencia. Pero está claro que las creencias accidentalmente erróneas también tienen causas, incluso pueden ser causadas por otros estados mentales (como las creencias resultantes de inferencias accidentalmente falaces). Lo que puede parecer arbitrario es que, según la definición de Fernández, los estados mentales que accidentalmente causan creencias no cuenten como sus fundamentos. Más bien, Fernández estipula con su definición que estas creencias no tienen fundamentos. Si, por el momento, aceptamos los estados mentales que causan accidentalmente a alguna creencia en el sujeto como "fundamentos irregulares", entonces el procedimiento de Bypass podría ser puesto en marcha también para conocer a estas creencias. El sujeto simplemente tendría que basar su creencia de segundo orden en el fundamento irregular de la creencia de primer orden. Parece que

Bypass debería funcionar igual que en el caso regular, excepto que la autoadscripción carecería de justificación. La situación entonces sería la siguiente: podríamos hacer autoadscripciones de creencias, tanto regularmente formadas como accidentalmente formadas, usando Bypass. En ambos casos la autoadscripción debería ser verdadera. Pero en el segundo caso le faltaría la justificación, porque la relación entre el fundamento (irregular) y la creencia de primer orden no sería regular. Y esta discrepancia entre las dos autoadscripciones es una consecuencia extraña de la teoría.

La segunda observación crítica es que incluso si aceptamos la estipulación de Fernández y aceptamos que las creencias accidentales no tienen fundamento y por ende no puede haber un procedimiento Bypass tal como Fernández lo describe para conocerlas, todavía queda la pregunta de si no es posible conocer tales creencias de la misma manera que creencias que no son erróneas de esta forma. Si mi creencia es el resultado de un error no habitual, ¿por qué significa eso que no puedo conocerla (de la manera normal en la que conozco todas mis creencias)? Fernández usa repetidamente el ejemplo donde veo una manzana en frente de mí y con base en esta percepción formo la creencia de segundo orden "Creo que hay una manzana en frente de mí". Pero supóngase que lo que veo en realidad es una pera. Normalmente la reconocería como tal. Pero en esta ocasión estoy distraído por alguna razón (el teléfono suena) y formo la creencia de que hay una manzana en frente de mí. Tal vez más tarde, reflexionando bien sobre lo que vi, concluya que probablemente era una pera (y podemos suponer que me baso exclusivamente en el recuerdo de mi percepción, no en información adicional que he recibido mientras). ¿Es plausible que anteriormente *no sabía* que creía que había una manzana en frente de mí? ¿Por qué no debería haber sido capaz de conocer esta creencia mía? Pero si pude conocerla, Bypass no nos explica cómo eso es posible.[28]

[28] Fernández sugiere que la objeción describe un caso muy raro, seriamente anormal: "Si las circunstancias psicológicas que estoy experimentando cuando formo la creencia accidental de primer orden son bastante anormales (y deben ser seriamente anormales si carezco de cualquier fundamento para mi creencia de primer orden), entonces no parece ser irrazonable pensar que, incluso si mi autoatribución de la creencia de primer orden es verdadera, no está justificada. Después de todo, si he tomado drogas o no he dormido por 48 horas y no soy responsable por la creencia de primer orden que estoy formando, ¿qué razón hay

La tercera observación crítica es de una naturaleza más general. Fernández afirma que la formación de creencias sobre fundamentos irregulares es algo muy raro que presupone condiciones seriamente anormales (cfr. Fernández, 2015b, p. 150). Pero —contrariamente a lo que piensa nuestro autor—, en general, errores accidentales no son raros, sino algo que ocurre cotidianamente. Tales errores, para Fernández, tienen que darse no en el paso de los estados mentales que son los fundamentos a las creencias de primer orden, sino anteriormente, en la formación de los estados mentales que —muchas veces— todavía no son creencias, sino percepciones, recuerdos o impresiones, y que sirven como fundamento para la formación de éstas. Así, según nuestro autor, es raro que algo falle en el paso de un estado perceptual a la creencia correspondiente o en el paso de una memoria aparente a la creencia correspondiente. Pero no es tan raro que algo falle en la formación del estado perceptual o en la formación de la memoria. Es porque Fernández ubica el error accidental principalmente en el proceso de la formación de los estados *anteriores* a la formación de creencias que su procedimiento de Bypass no parece ser afectado por tales errores. Pero ¿es plausible que la formación de creencias raramente falle y que la multitud de los pequeños o grandes errores que cotidianamente ocurren en nuestras creencias se deban principalmente a fallas anteriores a la formación de creencias?

Esta conclusión se podría cuestionar a través de una duda acerca de una afirmación general que Fernández hace: ¿es cierto que generalmente formamos nuestras creencias con base en otros *estados mentales*? Para Fernández, nuestras creencias sólo hacen contacto con la realidad a través de otros estados mentales. Pero si la realidad también pudiera llevar *directamente* a la formación de creencias, y si esto fuera un proceso muy frecuente, y si tal vez incluso fuera el proceso normal de formación de creencias, entonces sería más plausible que a menudo nuestra misma formación de creencias falle y produzca errores, no sólo la formación de estados mentales anteriores a la formación de creencias. En este caso podría haber un número significativo de creencias accidentalmente erróneas; creencias sin fundamento en otros estados mentales y formadas de manera irregular. Tales creencias ciertamente tendrían causas, pero estas causas no contarían como fundamentos, según la definición oficial

para pensar que soy competente en atribuir creencias a mí mismo?" (Fernández, 2015b, p. 150).

de fundamento (a saber, un estado mental que *tiende* a causar una creencia) que Fernández ofrece. Si pudiéramos *conocer* estas creencias no podría ser a través del procedimiento Bypass, ya que no tienen fundamento. Incluso si aceptáramos las causas irregulares que estas creencias indudablemente tendrían como sus fundamentos, es claro que un procedimiento como Bypass no podría proporcionarnos una justificación para nuestro conocimiento, ya que la relación entre estas causas y nuestras creencias sería irregular y accidental. Tendríamos, entonces, una situación en la que un número significativo de nuestras creencias —y no sólo creencias en situaciones raras— no podrían ser conocidas a través de Bypass o —si pudieran ser conocidas— no podrían ser justificadas a través de Bypass, porque se formarían erróneamente y de manera accidental e irregular.

Para evaluar esta crítica debemos preguntar si las creencias en realidad pueden basarse directamente en la realidad (afuera de nuestra mente), en lugar de a través de otros estados mentales intermediarios. Con respecto a la percepción, por ejemplo, varios filósofos argumentan que ésta no es un estado intermediario entre realidad y creencias, sino que percibir es sólo adquirir ciertas creencias (sobre cosas perceptibles).[29] Teorías similares existen sobre la memoria (recordar *es* directamente formar una creencia sobre el pasado, o recordar *es* todavía saber algo que ocurrió en el pasado,[30] no producir un estado mental que luego sirve de fundamento para formar la creencia). No es posible, en el contexto de este artículo, discutir estas teorías a fondo. Pero el hecho de que existen señala que la propuesta de Fernández puede ser cuestionada en este punto.

Para terminar esta línea argumentativa quisiera sólo mostrar un problema análogo en la manera en que Fernández entiende la adquisición de creencias por medio de inferencias. Dice el autor: "Si S cree que Q y S cree que P se sigue de Q, entonces S llega a creer que P" (Fernández, 2013, p. 46). Según este esquema, ¿cómo razonaría una persona que afirma el consecuente? Normalmente esta falacia se describe según el siguiente esquema:

---

[29] Cfr., por ejemplo, Armstrong (1968, capítulo 10).

[30] Gilbert Ryle parece entender así el concepto de "recordar" (cfr. Ryle, 1949, capítulo VIII, sección 7).

Si *p*, entonces *q*.
q
_____
Por lo tanto, *p*.

En el principio de formación de creencias que propone Fernández, en este caso Q representaría "(Si *p*, entonces *q*) y *q*" y P representaría *p*. Fernández dice que si un sujeto S cree que "(si *p*, entonces *q*) y *q*" y si S cree que *q* se sigue de "(si *p*, entonces *q*) y *q*", entonces S llega a creer que *p*. El problema con esta reconstrucción del razonamiento falaz es, a mi modo de ver, que la falacia *desaparece* del razonamiento. Si uno cree una cosa y cree que otra se sigue de esta cosa, entonces es racional llegar a creer esta segunda cosa. El problema de la falacia aquí se ubica únicamente en el valor de verdad de las premisas del razonamiento. Si no es cierto que la segunda cosa se sigue de la primera, entonces uno puede llegar a creer algo falso. Lo que falla es la creencia de que la segunda cosa se sigue de la primera. Esto no es una falla en la inferencia, sino en una de las premisas de la inferencia. Nuevamente resulta, entonces, que la formación de la creencia, tal como la concibe Fernández difícilmente puede fallar, mientras el error que contiene la creencia se debe a algo anterior a la formación. Sin embargo, si tomamos en serio la idea de que es posible *razonar falazmente*, entonces tenemos que decir que el sujeto cree que si *p*, entonces *q* y que *q*, y de esto *directamente* concluye que *p*. Es decir, las creencias en las que se basa el sujeto son la creencia de que si *p*, entonces *q* y la creencia de que *q*, no una creencia adicional de que de eso se sigue que *p*.[31] El razonamiento es falaz porque la conclusión no se sigue de las premisas tal como el sujeto razonó. La formación misma de

[31] La idea de que tal creencia podría ser necesaria para poder hacer la inferencia de hecho nos puede llevar a un regreso infinito. Fernández dice que el sujeto, para inferir de Q que P, debe creer que Q y creer que la conclusión P se sigue de la premisa Q. Pero ahora parece que hemos añadido una nueva premisa al argumento, a saber, que la conclusión P se sigue de la premisa Q. En lugar de una simple inferencia de Q a P, ahora tenemos una inferencia con forma de *modus ponens*. El regreso surge porque, siguiendo la pauta de Fernández, ahora se puede decir nuevamente que el sujeto debería creer una premisa adicional que dice que de las anteriores premisas (a saber: Q y que de Q se sigue P) se sigue la conclusión P. Y así al infinito. El regreso se puede evitar aceptando que un sujeto puede inferir de una premisa Q que P sin creer premisas adicionales.

la creencia falla, no sólo uno de los estados mentales (la creencia de que P se sigue de Q) en los que se basa como fundamento. La objeción que he discutido supone que la formación de la creencia a menudo no sólo es falaz, sino también irregular. Es plausible que aun así podamos tener un conocimiento peculiar y privilegiado de ella: ¿por qué no debería ser posible conocer una creencia accidental de la misma manera que una regular? Si es posible, la teoría de Fernández no puede explicar este tipo de autoconocimiento.

Veamos ahora mi segunda objeción a la teoría de Fernández, la cual concierne al aspecto metafísico de la teoría. Me parece que es la más seria: no es claro si la teoría puede explicar cómo un sujeto se puede dar cuenta de un *cambio* en la manera en que forma sus creencias de primer orden. Reconsideremos un ejemplo que anteriormente señalé como ejemplo de la formación irracional de creencias. Un sujeto razona falazmente afirmando el consecuente. Fernández argumenta que si el sujeto *regularmente* razona de esta manera falaz, entonces el procedimiento Bypass le puede dar un conocimiento de las creencias así formadas falazmente, porque la relación entre los fundamentos y las creencias de primer orden, aunque sea falaz, es regular. El sujeto *siempre* razona de esta manera y por eso está justificado cuando, sobre las mismas bases, también se autoatribuye la creencia que se formó de esta manera irracional. En la anterior objeción sobre el aspecto normativo de la teoría de Fernández discutimos esta idea. Aquí, quiero enfocarme en el aspecto metafísico. Supóngase que el sujeto *cambia* su manera de razonar. Deja de afirmar el consecuente y se conforma con afirmar el antecedente. ¿Puede el sujeto darse cuenta de este cambio en su manera de formar creencias?

La pregunta no es si el sujeto tiene alguna opinión sobre los procesos de su formación de creencias, sino simplemente si sus autoatribuciones de creencias, resultado de la aplicación del procedimiento Bypass, reflejarán el cambio en su manera de razonar. Según Fernández, Bypass significa que la creencia de segundo orden se forma únicamente sobre el fundamento de la creencia de primer orden, no con base en esta creencia misma. La dificultad, cuando cambia la manera de razonar del sujeto, es que, al parecer, el procedimiento de Bypass no puede ser sensitivo a tales cambios, ya que el procedimiento se basa únicamente

En este caso, lo que determina si el sujeto razona falazmente es la validez o invalidez del paso de Q a P, no si sus premisas son verdaderas o falsas.

en estados mentales que son anteriores a la formación de la creencia de primer orden. El cambio en la manera de formar creencias no parece ser detectable en los estados mentales a partir de los cuales luego se forman las creencias. En nuestro ejemplo, las creencias "*q*" y "Si *p*, entonces *q*" primero (cuando el sujeto todavía tiene el hábito de afirmar falazmente el consecuente) llevan a la formación de la creencia de que *p* y luego (cuando el sujeto deja el hábito de razonar así atrás) ya no. No hay nada en las creencias "*q*" y "Si *p*, entonces *q*" mismas que podría indicar si el sujeto las usa como fundamento para razonar (falazmente) que *p* o no. Entonces, ¿cómo se supone que el sujeto sepa —a través de Bypass solamente y con base en estas dos creencias— si cree que *p*?

Tal vez Fernández podría argumentar que la rutina de Bypass se puede formar con el tiempo porque existe una regularidad en la forma de razonar del sujeto. Primero existe un hábito de razonar falazmente y eso le permite al sujeto formar una rutina de Bypass correspondiente. Con el tiempo el sujeto aprende y se da cuenta de que autoatribuirse creencias según Bypass le resulta en autoadscripciones de creencia verdaderas. Igualmente, cuando el sujeto deja de razonar así, con el tiempo se va a dar cuenta de que la vieja rutina de Bypass ya no produce autoadscripciones verdaderas y en consecuencia la abandona. Vale mencionar que para así aprender por experiencia que cierto procedimiento de Bypass funciona, un sujeto tiene que conocer y confirmar sus creencias de primer orden de una manera independiente, probablemente desde la perspectiva de la tercera persona. No es claro que exista tal aprendizaje; si existiera, deberíamos estar familiarizados con él en nuestra vida diaria.

Pero me parece que la idea de Fernández es más bien que el sujeto puede saber qué creencia surgirá de sus estados de fundamento porque en virtud de estos estados *le parece* al sujeto que *p*. Cuando el sujeto percibe (aparentemente) una manzana, entonces al sujeto *le parece* que ahí hay una manzana. Cuando el sujeto recuerda (verídica o aparentemente) que *p*, entonces al sujeto *le parece* que *p*. Cuando el sujeto (que está acostumbrado a razonar falazmente) cree que *q* y que si *p*, entonces *q*, entonces —en virtud de estas creencias— *le parece* al sujeto que *p*. Y cuando el sujeto (que no tiene la disposición falaz) cree que *q* y que si *p*, entonces *q*, entonces —en virtud de estas creencias— *no le parece* al sujeto que *p*. Así, los estados mentales que son los fundamentos, según Fernández, para nuestras creencias de primer orden traen consigo algo como una tendencia o una inclinación a formar la creencia en cuestión. Y esta tendencia debe ser detectable por el sujeto, tal vez en la forma de

un "parecer", para que el procedimiento de Bypass funcione. Sólo si el sujeto de algún modo detecta que algún estado mental le mueve hacia la formación de cierta creencia (de primer orden) el procedimiento de Bypass se activa y produce la autoatribución de la creencia.

Si eso es así, entonces se trata de una adición importante a la teoría de Fernández. No es sólo la presencia de algún estado mental de fundamento lo que nos hace autoatribuir vía Bypass una creencia de primer orden (a saber, la creencia que normalmente formamos con base en el estado mental). Más bien, para que el procedimiento de Bypass funcione, el sujeto tiene que detectar además la fuerza que el estado mental ejerce sobre él para formar cierta creencia. Sólo gracias a detectar esta fuerza puede el sujeto seguir autoatribuyendo correctamente sus creencias incluso cuando cambia su manera de formarlas.

¿Es esta adición a la teoría de Fernández problemática? Creo que no; al contrario, hace la teoría más plausible. Pero la adición es significativa cuando se trata de comparar las teorías de Byrne y Fernández. Originalmente parecía haber una diferencia importante entre las dos que reside en lo que cada una considera como base para llegar a una autoadscripción de creencia. En el caso de Byrne, la autoadscripción se basa, vía una inferencia, en la creencia de primer orden misma que se autoadscribe. En cambio, Fernández sugiere que esta creencia de primer orden "se pasa por alto" en el procedimiento Bypass y la autoadscripción sólo se basa en lo que es el fundamento para la creencia de primer orden, no en esta creencia misma. Esta diferencia entre las dos teorías —la más importante a mi modo de ver— se reduce con la nueva adición a la teoría de Fernández. Según la adición, nuestras autoadscripciones se basan en los fundamentos para las creencias de primer orden *y* en la fuerza que estos fundamentos tienen para llevarnos a formar ciertas creencias de primer orden, aquéllas que luego autoadscribimos. Si comparamos lo que es la base para autoadscribir la creencia de que *p*, según Byrne y Fernández, tenemos lo siguiente: según Byrne, la base es la creencia de que *p*. Según Fernández, la base es un estado mental *E* y su actual fuerza para llevar al sujeto a formar la creencia de que *p*. Tal vez podríamos decir que para autoadscribir la creencia de que *p* tengo que detectar mi estado *E* y que en virtud de *E* parece que *p*.[32] Eso es muy similar a creer que

[32] Se trata de un parecer *tomando en cuenta toda la información que tengo*. En este sentido, un palo sumergido parcialmente en agua *no* parece doblado *tomando en cuenta que el agua distorsiona la visión* (cfr. Fernández, 2015b).

p. Y por eso las teorías de Byrne y Fernández se vuelven más similares una vez que la teoría de Fernández se modifica como aquí he indicado. Fernández requiere que el sujeto detecte que "parece que *p*" (en virtud de *E*) para luego autoatribuir la creencia de que *p*. Byrne requiere que el sujeto detecte que *p* (aunque es posible que se equivoque sobre *p* y sólo le parezca que *p*) para luego autoatribuir la creencia de que *p*. La diferencia principal entre las dos posiciones, según mi parecer, se reduce ahora a que según Fernández creer algo siempre sucede en virtud de otro estado mental *E*, mientras Byrne no se compromete con tal afirmación. Pero como vimos, esta afirmación puede ser cuestionada[33] y, además, ya no parece central para el procedimiento de Bypass. Lo central parece ser que el sujeto detecte que algo (el fundamento para su creencia de primer orden, sea esto un estado mental o no) lo "mueve" a creer algo; en otras palabras, el sujeto tiene que detectar que (parece que) *p*. Y como expuse en la sección 4, esto es también la base del procedimiento que Byrne describe con su regla Cree.

8. Conclusión

Tanto Byrne como Fernández ilustran cómo una teoría de la transparencia de la mente puede ayudarnos a solucionar el problema del autoconocimiento, es decir a explicar y reconciliar tanto la peculiaridad como el privilegio que acompañan a este fenómeno. La peculiaridad se debe a que sólo en mi propio caso tiene sentido utilizar un procedimiento de transparencia para autoatribuir creencias; lo que se me presenta como verdadero sobre el mundo es precisamente lo que *creo* sobre él, aunque no necesariamente lo que cree alguien más sobre él. El privilegio (el cual no incluye infalibilidad, según esta propuesta) se debe a la sencillez del proceso de la autoatribución, la cual no depende de la veracidad de la percepción, memoria, testimonio u otras fuentes del conocimiento tales como inferencias complejas.

En la discusión de las dos propuestas de transparencia vimos que ambas se enfrentan a ciertas objeciones: la racionalidad del procedimiento puede ser cuestionada (cfr. Boyle, 2011, y mi primera

[33] De hecho, Byrne parece rechazar la afirmación para el caso de creencias perceptuales. Según él, percibir involucra constitutivamente formar creencias perceptuales de tal forma que la percepción no puede ser un estado mental meramente intermediario entre realidad y creencia (cfr. Byrne, 2012a, pp. 204 y ss.).

objeción a Fernández) y también su funcionalidad (cfr. Gertler, 2011, y mi segunda objeción a Fernández). Tanto Byrne como Fernández pueden contestar estas objeciones. Pero me parece que en el caso de Fernández los problemas son un poco más graves y, cuando tratamos de solucionarlos, su teoría se asemeja mucho a la de Byrne. Quedan otros aspectos de las teorías para investigar, en especial cómo se comparan los intentos de ambos filósofos por generalizar sus procedimientos para la adquisición no sólo del conocimiento de las propias creencias, sino también de otros estados mentales, como los deseos.[34] El resultado del presente trabajo es que la propuesta de Byrne parece ser algo más plausible que la de Fernández.[35]

Referencias

Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. Routledge & Kegan Paul.

Boyle, M. (2011). Transparent Self-Knowledge. *Proceedings of the Aristotelian Society Supplementary*, 85(1), 223-241. https://doi.org/10.1111/j.1467-8349.2011.00204.x.

Byrne, A. (2005). Introspection. *Philosophical Topics*, 33(1), 79-104. https://doi.org/10.5840/philtopics20053312.

___ (2011a). Knowing That I Am Thinking. En A. Hatzimoysis (ed.), *Self-Knowledge*. (pp. 105-123). Oxford University Press.

___ (2011b). Transparency, Belief, Intention. *Proceedings of the Aristotelian Society Supplementary*, 85(1), 201-221. https://doi.org/10.1111/j.1467-8349.2011.00203.x.

[34] El libro de Fernández elabora en detalle cómo un procedimiento análogo a Bypass podría permitirnos autoatribuir deseos (cfr. Fernández, 2013, capítulo 3). El autor también muestra cómo su modelo podría ayudarnos a entender la paradoja de Moore, el fenómeno de la inserción de pensamientos y la autodecepción (cfr. Fernández, 2013, capítulos 4, 5 y 6). Byrne, por su parte, ha publicado textos que generalizan la idea detrás de la regla Cree al conocimiento de los propios deseos (Byrne, 2012b), las propias intenciones (Byrne, 2011b), las propias percepciones (Byrne, 2012a) y los propios pensamientos actuales (Byrne, 2011a).

[35] Agradezco a dos árbitros de *Tópicos* por sus comentarios y sugerencias acertadas para mejorar este trabajo.

___ (2012a). Knowing What I See. En D. Smithies y D. Stoljar (eds.), *Introspection and Consciousness*. (pp. 183-209). Oxford University Press.

___ (2012b). Knowing What I Want. En J. Liu y J. Perry (eds.), *Consciousness and the Self: New Essays*. (pp. 165-183). Cambridge University Press.

Descartes, R. (1904). Œuvre *de Descartes. Vol. VII. Meditationes de prima philosophia*. C. Adam y P. Tannery (eds.). Léopold Cerf. Sitio de Internet Archive: https://archive.org/details/uvresdedescartes07desc/page/n8.

Evans, G. (1982). *The Varieties of Reference.* Clarendon.

Fernández, J. (2013). *Transparent Minds: A Study of Self-Knowledge.* Oxford University Press.

___ (2015a). Resumen de *Transparent Minds*. *Teorema*, 34(1), 95-100. Sitio de Dialnet: https://dialnet.unirioja.es/servlet/articulo?codigo=4994446.

___ (2015b). Replies to my Critics. *Teorema*, 34(1), 149-160. Sitio de Dialnet: https://dialnet.unirioja.es/servlet/articulo?codigo=4994450.

Fricke, M. (2009). Evans and First Person Authority. *Abstracta. Linguagem, Mente & Ação*, 5(1), 3-15. Sitio de Abstracta: http://abstracta.oa.hhu.de/index.php/abstracta/article/view/109.

_____(2012). Racionalidad y autoconocimiento en Shoemaker. En P. Stepanenko (ed.), *La primera persona y sus percepciones.* (pp. 53-73). UNAM, CEPHCIS.

Gallois, A. (1996). *The World Without, the Mind Within: An Essay on First-Person Authority.* Cambridge University Press.

Gertler, B. (2011). Self-Knowledge and the Transparency of Belief. En A. Hatzimoysis (ed.), *Self-Knowledge*. (pp. 125-145). Oxford University Press.

Moran, R. (2001). *Authority and Estrangement. An Essay on Self-Knowledge.* Princeton University Press.

___ (2003). Responses to O'Brien and Shoemaker. *European Journal of Philosophy*, 11(3), 402-419. https://doi.org/10.1111/1468-0378.00193.

Parrott, M. (2015). Review of *Transparent Minds: A Study of Self-Knowledge*, by Jordi Fernández. *European Journal of Philosophy*, 23(S1, Reviews Supplement), e19-e22. https://doi.org/10.1111/ejop.12115.

Ryle, G. (1949). *The Concept of Mind.* Hutchinson.

Shoemaker, S. (1994). Self-Knowledge and 'Inner Sense'. *Philosophy and Phenomenological Research*, 54(2), 249-314. https://doi.org/10.2307/2108488, https://doi.org/10.2307/2108489, https://doi.org/10.2307/2108490.

Thomas, J. (2014). *The Minds of the Moderns. Rationalism, Empiricism and Philosophy of Mind*. Routledge.

Tugenhat, E. (1993). *Autoconciencia y autodeterminación: Una interpretación lingüístico-analítica*. Fondo de Cultura Económica.

Valaris, M. (2011). Transparency as Inference: Reply to Alex Byrne. *Proceedings of the Aristotelian Society*, 111(2), 319-324. https://doi.org/10.1111/j.1467-9264.2011.00312.x.