CRÍTICA, Revista Hispanoamericana de Filosofía. Vol. 49, No. 146 (agosto 2017): 37–60

HAY MUCHAS COSAS QUE CREO DE MÍ MISMO (CONSCIENTE E INCONSCIENTEMENTE) SIN SABER QUE LAS CREO

MIGUEL ÁNGEL SEBASTIÁN
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
msebastian@gmail.com

RESUMEN: En un artículo publicado recientemente (2014) en esta revista, Javier Vidal argumenta que toda creencia de primera persona es una creencia consciente, una conclusión que pone en jaque ciertas teorías de la consciencia, como él mismo expone. El razonamiento de Vidal se basa en un argumento que muestra que uno conoce toda creencia de primera persona que tiene y en un principio (SC*) que vincula conocimiento y consciencia. Mi objetivo en este trabajo es mostrar que el razonamiento de Vidal no es sólido. En particular, hago patente que el argumento depende de rechazar la relación ampliamente aceptada en epistemología entre creencia y conocimiento. Además, argumento que SC* o bien prejuzga la cuestión o bien involucra una noción de consciencia no relevante para la discusión.

PALABRAS CLAVE: consciencia, representación *de se*, teorías de pensamiento de orden superior, creencia inconsciente, referencia de primera persona

SUMMARY: In a recent paper in this journal (2014), Javier Vidal has argued that every first-person belief is a conscious one, a conclusion that jeopardizes certain theories of consciousness as he shows. Vidal's reasoning is builded upon an argument to the effect that one knows all first person beliefs that one has and a principle (SC*) that links knowledge and consciousness. My aim in this paper is to show that Vidal's reasoning is unsound. In particular, I show that the argument depends upon the rejection of the relation, widely accepted in epistemology, between belief and knowledge. Moreover, I argue that SC* either begs the question or involves a notion of consciousness that is not relevant for the discussion.

KEY WORDS: consciousness, *de se* representation, higher-order thought theories, unconscious belief, first-person reference

Los estados cuyo contenido es de primera persona parecen desempeñar un papel fundamental en la explicación del comportamiento y la agencia. Así mismo, varios autores han señalado que la consciencia ha de ser entendida, al menos a cierto nivel, en términos de estados con contenidos de primera persona. Javier Vidal (2014, 2015) ha argumentado que, necesariamente, toda creencia de primera persona ha de ser una creencia consciente. Este interesante resultado —extrapolable a otras actitudes proposicionales como los deseos, las dudas, etc.— no sólo pone en jaque cierto tipo de explicaciones de la consciencia, como el propio Vidal señala, sino que además presenta un problema para aquellas teorías que explican ciertos aspectos del

comportamiento postulando estados mentales inconscientes con contenido de primera persona. El objetivo de este artículo es mostrar que el argumento de Vidal no es sólido y que no existe el tipo de conexión entre consciencia y creencias de primera persona que él pretende establecer.

Este artículo está organizado de la siguiente manera: en la primera sección, presento los contenidos de primera persona y su conexión con la explicación del comportamiento y de la consciencia. En la segunda, expongo detalladamente el argumento en favor de que las creencias de primera persona han de ser necesariamente conscientes. En la última sección, examino críticamente el argumento y muestro que las condiciones bajo las cuales una de las premisas podría resultar aceptable son tales que las otras premisas resultan falsas. Concluyo que el argumento es falaz y que las teorías consideradas no tienen nada que temer del argumento de Vidal.

1. *El problema de las creencias* de se

Los seres humanos podemos tener distintos tipos de actitudes proposicionales —como creer, desear o esperar— hacia distintas cosas. Uno puede creer, por ejemplo, que las películas de luchadores son estupendas o que el chocolate es bueno para la salud. Decimos que estos estados mentales son *estados representacionales* o que tienen contenido. Una pregunta que surge inmediatamente al tratar de entender la naturaleza de este tipo de estados es cuál es su contenido.

La respuesta tradicional a esta pregunta semántica sostiene que las *proposiciones* son el contenido de nuestras actitudes. Pese al notable desacuerdo acerca de la naturaleza de las proposiciones,[1] está ampliamente aceptado que son el tipo de cosa cuya adecuación o veracidad es evaluable dependiendo de cómo sea el mundo (dejando abierta la posibilidad de que una proposición pueda carecer de valor de verdad). En otras palabras, las proposiciones hacen particiones entre mundos posibles en el espacio de posibilidad; *i.e.* determinan

[1] Algunos (Stalnaker 1999) piensan que las proposiciones son meros conjuntos de mundos posibles. Los que se oponen a esta posición mantienen que las proposiciones tienen una estructura que es básicamente análoga a la de una oración. Por ejemplo, la proposición *que las películas de luchadores son estupendas* se puede identificar con una estructura como <películas de luchadores, ser estupendas>. La visión estructuralista puede ser dividida entre teorías *russellianas* y *fregeanas*. La primera defiende que las proposiciones están constituidas de relaciones y particulares, mientras que la segunda acepta sentidos fregeanos (conceptos, modos de presentación) como elementos constituyentes de la proposición. Para una discusión reciente acerca de la naturaleza de las proposiciones véase King, Soames y Speaks 2014.

colecciones de mundos posibles. Por ejemplo, la proposición *que la Ciudad de México es la capital de México* es verdadera en algunos mundos, entre los cuales está el mundo actual, y falsa en otros. De forma más importante, comúnmente se considera que el valor de verdad de una proposición no cambia dependiendo del sujeto, el lugar o el tiempo. Las creencias que tenemos *acerca de nosotros mismos como tales* (Castañeda 1966; Chisholm 1981) parecen particulares en este sentido. Veámoslo con un ejemplo.

El Santo, Rodolfo Guzmán Huerta, es probablemente el luchador enmascarado más famoso del mundo. Se convirtió en un héroe popular por sus apariciones en libros de cómics y películas, y su máscara plateada es un símbolo de la justicia. Consideremos dos situaciones hipotéticas en la vida de El Santo:

> 1. Tras su retiro, El Santo está viendo *Las momias de Guanajuato*, donde Blue Demon, Mil Máscaras y él luchan contra un grupo de momias reanimadas. Fascinado por sus propias habilidades en la lucha piensa: Yo era un gran luchador.
>
> 2. Tras su retiro, El Santo tiene un accidente de tráfico y, debido a un golpe en la cabeza, sufre amnesia postraumática y no puede recordar que llevaba una máscara y que era un famoso luchador. Un día, mientras se recupera, está viendo *Las momias de Guanajuato*, donde Blue Demon, Mil Máscaras y él luchan contra un grupo de momias reanimadas. Fascinado por sus propias habilidades en la lucha piensa: El Santo era un gran luchador.

Ambas situaciones involucran creencias acerca de El Santo, pero hay una diferencia notoria: (1) pero no (2) involucra una "creencia de primera persona", una creencia *de se*. Mientras que en (1) El Santo tiene la creencia de que *él mismo* (Castañeda 1966) es un gran luchador, en (2) sólo tiene una "creencia de tercera persona". En esta última situación, El Santo falla en darse cuenta de que él es El Santo y no cree de él mismo que fuera un gran luchador. Además, no se da cuenta de que él es el luchador enmascarado al que está viendo en la pantalla. Parece que, sin importar qué propiedad *F* consideremos, El Santo podría creer que *F* es un gran luchador, sin creer que él mismo es el único individuo que satisface *F*, por lo tanto, sin poseer la correspondiente creencia *de se*.[2] Las actitudes de primera persona

[2] Como el propio Vidal señala (2014, nota 3), la locución 'él mismo' o 'ella misma' no siempre es utilizada para reportar una actitud *de se* y además, a menudo,

o actitudes *de se* parecen ser actitudes *sui generis*,[3] no reducibles a otro tipo de actitudes proposicionales.

Lo que demandan del mundo tanto la creencia de primera persona como la de tercera para ser verdaderas es exactamente lo mismo: que El Santo fuera un gran luchador. Sin embargo, son creencias de distinto tipo y un sujeto puede tener una sin tener la otra. Diversos autores han tratado de capturar estas diferencias de distinto modo. Por ejemplo, Perry (1979) considera que el contenido es el mismo y trata de capturar las diferencias en términos del tipo de estado involucrado. En una línea similar, algunos autores defienden que las correspondientes creencias de primera y tercera personas no difieren en la proposición involucrada, sino en el modo de presentación o *guisa* en la que es creída esa proposición (Ezcurdia 2001, Kaplan 1983, Perry 1980, Richard 1983, Salmon 1981). Alternativamente, Lewis (1979) consideró que la diferencia ha de ser explicada a nivel de contenido. Lewis argumentó que lo que los ejemplos muestran es que el tipo de particiones que hacen las creencias de primera persona es más fina que aquellas que hacen las proposiciones (entendidas en la forma en que las he presentado): las creencias de primera persona no son verdaderas o falsas dependiendo únicamente de cómo sea el mundo, sino también del sujeto que tiene la creencia. Si hay diferencias sustantivas entre estas posiciones o si son meramente verbales, es una discusión abierta y, sobre esto, pretendo permanecer neutral en el resto del texto.[4]

Las actitudes proposicionales se postulan, entre otras cosas, para explicar nuestro comportamiento. Mi deseo de tomar una cerveza y mi creencia de que hay cervezas en el refrigerador explican, al menos parcialmente, el hecho de que me levante y me dirija al frigorífico. Explicar no sólo algunos, sino todos los comportamientos requiere apelar a creencias de primera persona como muestra Perry (1979).[5]

Imaginemos que El Santo está en el supermercado comprando algunos productos cuando ve a un ladrón armado entrar y amenazar a los clientes. Instantes después, el ladrón resbala con azúcar derramada en el suelo, cae y queda inconsciente. El Santo sigue el rastro de azúcar en el suelo del supermercado, empujando su carrito en busca

las adscripciones *de se* se construyen en infinitivo con un sujeto implícito o sobre entendido como "*X* cree estar Φ-endo". Por ello, la expresión pronominal PRO podría ser más adecuada para desempeñar ese papel (para una discusión más detallada, véanse Dowty 1985, Cappelen y Dever 2013, Chierchia 1989, y Corazza 2004).

[3] *Cfr.* Cappelen y Dever 2013.

[4] Para una revisión, véase Feit 2008.

[5] *Cfr.* Cappelen y Dever 2013.

del cliente que la había derramado, para felicitarlo por ser un héroe. Finalmente, El Santo se da cuenta de que es él mismo quien había estado derramando el azúcar: él era el héroe al que quería felicitar.

Parece haber dos tipos distintos de creencia involucrados en esta historia: la que tenía El Santo antes de percatarse de que él era el que estaba derramando el azúcar y la que tiene después. El Santo creía que el cliente que derramó el azúcar era un héroe pero no que él mismo era un héroe —al menos no por los acontecimientos de esta historia particular—. Cuando adquiere la nueva creencia cambia su comportamiento, por ejemplo, recolocando el paquete de azúcar y dejando de buscar al heroico cliente. Los ejemplos de Perry (1979) (o Lewis 1979) muestran una conexión esencial entre las creencias de primera persona y la acción que resulta de ella. Una conexión que ha sido ampliamente aceptada y rara vez discutida.[6]

Las creencias de primera persona también se han utilizado para ofrecer una teoría de la consciencia en términos de contenido de primera persona. La idea es que al tener un experiencia consciente uno no sólo se percata de que tal y cual es el caso, sino también, en uno u otro sentido, de que uno mismo está teniendo la experiencia. Este truismo, conocido como "principio de transitividad", da soporte a las teorías de orden superior. Estas teorías tratan de explicar la diferencia entre los estados que son conscientes y los que no los son en términos de otra relación que se da entre el estado consciente y algún tipo de representación de orden superior del anterior. En el caso particular de las teorías de pensamiento superior (Gennaro 1996, 2012; Rosenthal 1997, 2005), la representación de orden superior tiene forma de pensamiento de primera persona. Cuando tengo una experiencia como de rojo me encuentro en un estado con cierto contenido; llamemos a este contenido ROJO. Para que este estado mental sea consciente ha de haber adicionalmente un pensamiento de orden superior apuntándole cuyo contenido es algo así como 'Yo veo ROJO' (Rosenthal 1997).[7] Según la teoría, no hay necesidad de

[6] Véase Cappelen y Dever 2013, especialmente el capítulo tres, para una revisión crítica contraria a la ortodoxia de la conexión entre creencias de primera persona y acción.

[7] Algunos autores (Kriegel 2009, Sebastián 2012a) han tratado de capturar este principio de transitividad sin apelar a pensamientos de orden superior. La teoría que propongo (Sebastián 2012a), si bien no depende de pensamientos, sí trata de explicar la diferencia entre los estados conscientes y los que no lo son en términos de contenidos de primera persona elaborando la idea lewisiana de autoatribución. Por su lado, la teoría propuesta por Kriegel tampoco apela a pensamientos, pero no está claro que sea capaz de capturar el aspecto de primera persona que subyace al principio de transitividad (Prinz 2012, Sebastián 2012b).

que el estado de orden superior sea consciente y, de hecho, típicamente no lo es. Si la teoría requiriese que el estado de segundo orden fuera consciente, estaría en graves problemas pues, según la teoría, ello requeriría un estado de tercer orden que a su vez debería ser igualmente consciente por lo que se requeriría, a su vez, la presencia de un estado de cuarto orden y así *ad infinitum*. Por lo tanto, según este tipo de teorías, nuestra capacidad de tener experiencias conscientes depende de nuestra capacidad de tener pensamientos de primera persona que no son conscientes.

En el artículo publicado en *Crítica* (2014, no. 138), Javier Vidal argumenta que las creencias de primera persona son necesariamente conscientes. Si su argumento es sólido, el resultado pondría en jaque —como él mismo señala (*cfr.* Vidal 2014)— toda una familia de teorías de la consciencia que, como hemos visto, hacen depender la diferencia entre estados conscientes e inconscientes de contenidos de primera persona que puedan ser inconscientes.[8] Además, si la explicación del comportamiento requiere, al menos en algunos casos, tener estados con contenido de primera persona y éstos son necesariamente conscientes, entonces cierto tipo de explicaciones en términos de creencias y deseos de primera persona inconscientes resultarían directamente falsos al no existir ese tipo de estados. Debido las importantes implicaciones que parece tener el resultado del argumento de Vidal, amerita ser evaluado con detenimiento. En este artículo pretendo mostrar que o bien el argumento de Vidal prejuzga la cuestión simplemente aseverando aquello que pretende mostrar, o

[8] Vidal (2015) ofrece una posible salida al defensor de las teorías de orden superior. Valiéndose de la distinción de Recanati (2007, 2012) entre representación implícita y explícita argumenta que las representaciones *de se* implícitas no tienen por qué ser conscientes, por lo que el defensor de la teoría HOT podría apelar a ellas para salvar la teoría. Esta solución no parece, sin embargo, satisfactoria. Según Recanati, en la representación implícita el sujeto mismo es parte de las condiciones de corrección pero no queda determinado por el contenido sino por el modo de representación. Es este último el que garantiza que aquello que permanece fijo e invariable no sea parte del contenido pese a serlo de las condiciones de corrección. Para Recanati, el modo que garantiza tal cosa es el modo experiencial que, según él, es intrínsecamente de primera persona (2012, p. 185). En la teoría de Recanati la experiencia consciente fundamenta las representaciones implícitas de primera persona. Las teorías HOT buscan justo lo contrario, fundamentar la experiencia consciente en ciertas representaciones de primera persona. Para abrazar la propuesta de Vidal haría falta una teoría del modo de representación que sea independiente de la experiencia y que garantice que el estado no puede ser evaluado más que con respecto al sujeto. Vidal llama a ese estado modo interno pero no dice nada acerca de su naturaleza que vaya más allá de lo señalado por Recanati (incluso parece abrazar la teoría de éste; 2015, p. 130) y no conozco teoría alguna alterna.

bien, sus premisas no pueden ser verdaderas de forma simultánea: las consideraciones que podrían hacer una de las premisas verdadera hacen que otra de las premisas sea falsa.

2. *¿Son las creencias* de se *conscientes? El argumento*

El argumento de Vidal tiene dos partes. En la primera, ofrece un argumento del que concluye que si un sujeto tiene una creencia consciente acerca de sí mismo como tal, éste sabe que la tiene. En la segunda parte, se apela a un principio que nos permita ir desde la conexión entre cierto tipo de conocimiento y consciencia hasta la conclusión de que, necesariamente, toda creencia de primera persona es consciente. Veámoslo en detalle.

Consideremos la creencia que El Santo tiene en la situación (1) en el ejemplo anterior, cuyo contenido sería *que yo era un gran luchador*. En este caso la atribución de creencia sería como sigue: El Santo cree que él mismo era un gran luchador. Cuando uno tiene una creencia de primera persona —esto es, una creencia que refiere al sujeto de la creencia como tal—, como la creencia de El Santo que estamos considerando, la referencia de primera persona está garantizada. Parece que cuando alguien cree algo de sí mismo, como tal emplea el pronombre 'yo' o su análogo mental, y la referencia, por tanto, necesariamente existe pues es aquel que está teniendo la creencia.

Shoemaker (1968), siguiendo la distinción de Wittgenstein de dos usos de 'yo', distingue dos formas en las que uno puede representarse a sí mismo o dos tipos de creencias que incluyan el 'yo'. Consideremos la creencia de que tengo dolor dental formada sobre la base de mi dolor, y la creencia de que tengo un brazo enyesado que me formo tras ver un reflejo en el espejo. Esta última, a diferencia de la primera, no es *inmune al error por fallo en la identificación de la primera persona* (IEM), pues involucra reconocer un objeto particular, a una persona, y por tanto la posibilidad de error en tal reconocimiento. En el primer caso "no hace falta ningún reconocimiento cuando digo que yo tengo dolor dental. Preguntar ¿estás seguro de que eres tú el que tiene dolor? carecería de sentido" (Wittgenstein 1958, pp. 66–67).

Aun en los casos en que no hay IEM, resulta muy plausible pensar que hay un sentido en el cual el sujeto no puede equivocarse y al cual podemos llamar *inmunidad por fallo en la referencia* (IER). Éste es el sentido en el que, al tener una representación *de se*, con independencia de si presentan IEM o no, el sujeto no puede equivocarse en que aquel a quien refiere el pronombre es el mismo

que tiene la representación (O'Brien 2007; Rovane 1987). Este tipo de inmunidad es perfectamente capturada por Rosenthal:

> Yo veo el reflejo de alguien en un espejo y pienso erróneamente que yo soy esa persona [...]. Si considero que la persona en el espejo soy yo, puedo estar equivocado acerca de si el reflejo es realmente mío. Pero aun aquí no puedo estar equivocado [en cuanto a] de quién yo considero que es el reflejo; considero que el reflejo es del mismo individuo que está haciendo esa consideración. (2004, p. 173; 2005, p. 356, citado por Vidal 2014, p. 41.)

En palabras de Vidal:

> Como en esa creencia [la persona] se refiere con éxito a ella misma como *ella misma*, es imposible que la persona referida por ["ella misma"] en la adscripción (fuera de la cláusula-que) [...] ignore la identidad que guarda con la persona referida por "ella misma" en la adscripción (dentro de la cláusula-que), quien es el objeto de la creencia. (2014, p. 57)

Con la aceptación de IER, Vidal se permite aseverar que El Santo sabe que la creencia de que alguien fue un gran luchador es acerca de la misma persona que tiene la creencia de que alguien fue un gran luchador, esto es, él mismo. Al tener la creencia de primera persona en cuestión, El Santo no puede creer que aquel de quien cree que fue un gran luchador y aquel que tiene la creencia sean distintos. De ello Vidal concluye que El Santo sabe que él mismo cree que él mismo fue un gran luchador. Así, partiendo de una creencia de primera persona que un sujeto tiene, Vidal muestra que podemos concluir que el sujeto que tiene tal creencia sabe que la tiene. Podemos ahora reconstruir esta primera parte del argumento, a la que me referiré como CONOCERME, de forma un poco más formal.

CONOCERME

> Si un sujeto X cree que él mismo Φ, la siguiente aseveración desde la primera persona es verdadera:

1. Yo creo que yo Φ.

La atribución de creencia desde la tercera persona sería:

2. X cree que ella misma Φ.

Por IER, como hemos visto en el ejemplo de El Santo:

3. X sabe que la creencia de que alguien Φ es acerca de quien cree que alguien Φ, es decir, ella misma.

De ahí podemos concluir que

4. X sabe que ella misma cree que ella misma Φ,

como queríamos demostrar.

Una vez que se ha establecido que tengo conocimiento de mis creencias de primera persona, lo que queda para establecer que tales creencias han de ser conscientes es un vínculo entre conocimiento y consciencia. Vidal presenta un principio que parece plausible:

> SC: Si X sabe que ella misma cree que α, entonces X cree conscientemente que α.

Sin embargo, Vidal rechaza este principio dada la existencia de creencias inconscientes.

Siguiendo a Moran (2001) y Finkelstein (2003), Vidal presenta un contraejemplo a SC considerando el conocimiento que una persona pudiera tener tras una sesión de psicoanálisis de su creencia inconsciente de que comer es obsceno. Vidal señala que:

> aunque esa persona sabe ahora que ella misma cree que comer es obsceno y, por tanto, en cierto sentido tiene ahora conciencia *de* su creencia o es consciente de *que* cree que comer es obsceno, con todo, no cree conscient*emente* que comer es obsceno. Puede decirse que la persona psicoanalizada ahora tiene conocimiento, o conciencia, de su creencia *inconsciente* de que comer es obsceno. (p. 44, las cursivas son del original, y el subrayado mío.)

En este caso, aun cuando la persona esté dispuesta a autoadscribirse la creencia de que comer es obsceno y aseverar de forma sincera "yo creo que comer es obsceno", no está dispuesta a aseverar "comer es obsceno" y, por tanto, esta última creencia no es consciente en el sentido relevante para el argumento. Tras rechazar SC, Vidal indaga en las condiciones bajo las cuales podemos establecer una conexión entre conocimiento y consciencia y postula el siguiente principio:

> SC*: Si (si X cree que α, entonces X sabe que ella misma cree que α), entonces (si X cree que α, entonces X cree conscientemente que α).

El ejemplo mencionado anteriormente no constituye un contraejemplo a SC* ya que el ejemplo no hace verdadero el antecedente.[9] Si el principio es verdadero y CONOCERME es un argumento sólido, entonces, podemos establecer que toda creencia de primera persona es consciente.

En la siguiente sección mostraré que CONOCERME no es un argumento sólido, en particular, que de la inmunidad por fallo en la referencia no se sigue que haya conocimiento si aceptamos, como se hace ampliamente en epistemología, que saber implica creer. Exploraré diversas formas de establecer esa conexión y argumentaré que rescatar la vinculación entre las premisas 2 y 3 deja a SC* sin justificación.

3. *Podría creer cosas de mí mismo sin saber que las creo*

En esta sección pretendo demostrar que el argumento de Vidal no es sólido. En una primera subsección examino la relevancia del resultado del argumento al concluir una tesis según la cual las creencias de primera persona gozarían de un estatus epistémico probablemente único entre nuestros estados mentales que además lo pone en tensión con los argumentos presentados por Williamson en contra de la *luminosidad* de nuestros estados mentales. En la segunda subsección analizo detalladamente el argumento. Como hemos visto, el argumento de Vidal tiene dos partes. En la primera presenta un argumento CONOCERME que muestra que si un sujeto tiene una creencia de primera persona entonces sabe que la tiene. En una segunda parte, Vidal busca un principio plausible que conecte ese tipo de conocimiento con la consciencia. Lo que haré en esta sección es mostrar que el argumento del que concluye que siempre que tenemos una creencia consciente sabemos que la tenemos resulta inválido bajo el supuesto ampliamente aceptado en epistemología: saber implica creer. Más aún, Vidal argumenta que el principio que presenta en la segunda parte no sólo es plausible sino además *a priori*. Sin embargo, no ofrece ningún argumento a su favor más allá de sus intuiciones. Tras analizar la noción de consciencia relevante en el debate en que

[9] Nótese que cualquier deseo de primera persona inconsciente parecería constituir un contraejemplo a SC* (el argumento es independiente de la actitud proposicional como el propio Vidal (2014, nota 2) señala). Tras visitar al psicoanalista puedo descubrir que tengo el deseo freudiano hacia mi madre, un deseo de primera persona que, según el argumento, de existir habría de ser consciente. En respuesta, uno puede argumentar que tal aseveración prejuzga la cuestión en contra del principio, ya que si el argumento es válido, no hay creencias y deseos de primera persona, pero ello requeriría que hubiera un argumento como veremos en la sección siguiente.

participan las teorías de orden superior, y por tanto la que debería ser relevante en el argumento, mostraré que tal intuición no tiene soporte. Por esta razón, exploro la posibilidad de encontrar un argumento que pudiera dar sustento al principio que vincula cierto tipo de conocimiento y consciencia, y encuentro uno. Empero, este argumento es inconsistente con el argumento presentado en la primera parte que permite concluir que cuando tengo una creencia acerca de mí mismo sé que la tengo y, por tanto, nos deja sin razones para creer que toda creencia de primera persona es consciente.

3.1. Luminosidad y conocimiento

CONOCERME pretende mostrar que si X cree que ella misma Φ, entonces X sabe que tiene esa creencia; esto es, X sabe que ella misma cree que ella misma Φ. Ello implica que las creencias de primera persona son *luminosas* (Williamson 2000), donde un estado M de un sujeto S es luminoso si y sólo si, si M es el caso, entonces S está en posición de saber que M es el caso (Williamson distingue entre estar en posición de conocer y conocer. Si S está en posición de conocer que M es el caso y profundamente considera si M es el caso, entonces S conoce que M es el caso). De hecho, Vidal argumenta algo aún más demandante que la luminosidad de las creencias de primera persona, algo que podemos llamar 'hiperluminosidad'. Elaboraré con detalle esta idea para mostrar lo fuerte que resulta la conclusión de su argumento.

Sea L una relación de dos términos tal que '$L(M, S)$' quiere decir que M es luminoso para S; podemos definir de forma análoga 'K (S, M)' como S sabe que M es el caso, 'B (S, M)' como S cree que M es el caso y '$\blacklozenge K$ (S, M)' como S está en posición de saber que M es el caso.

Por definición, si un estado M es luminoso, entonces, uno está en posición de conocerlo (M & $L(M, S) \rightarrow \blacklozenge K$ (S, M)). Pero uno puede estar en posición de conocer que M es el caso y de hecho no conocerlo ($\neg$ ($\blacklozenge K$ (S, M)$\rightarrow$ K (S, M))). Ni siquiera el hecho de que el sujeto se forme la creencia acerca de que un estado luminoso es el caso basta para garantizar el conocimiento ($\neg$ (M & $L(M, S)$ & B (S, M))$\rightarrow$ K (S, M)). En general, aun cuando se esté en posición de saber p, uno puede formarse la creencia por razones incorrectas y, por lo tanto, no tener conocimiento. Imaginemos a dos jueces que tienen toda la evidencia para saber que Luisa robó el banco. Uno de ellos se forma la creencia por consideración de tal evidencia, mientras que el otro se la forma por leer los posos del café. Diríamos que el

primero sabe que Luisa robó el banco, mientras que el segundo no (Turri 2010). Igualmente, supongamos que *M* es luminoso pero *S* cree que *M* es el caso como consecuencia de lo que le revelan los posos del café. Aquí no diríamos que *S* sabe que *M* es el caso pese a que, de hecho, *M* sea el caso y *S* esté en posición de saberlo. Lo que Vidal exige, por tanto, es una condición a la que podemos llamar '*hiperluminosidad*' ($H(M, S)$) según la cual si *M* es el caso y *M* es hiperluminoso, entonces *S* sabe que *M* es el caso ($H(M, S) \rightarrow K(S, M)$). Ello implica que, o bien el conocimiento no está mediado por una creencia, o bien está mediado por una creencia que no puede formarse por las razones no adecuadas y que de hecho se forma con independencia si el sujeto considera a fondo si *M* es el caso. Parece perfectamente coherente que la creencia de que yo mismo tengo una creencia de primera persona, cuando de hecho la tengo, se pueda formar de forma inadecuada y por tanto no dar lugar a conocimiento, como ocurre en el ejemplo de los posos de café. No se me ocurre ninguna razón por la cual esto no pueda ser así y Vidal tampoco ofrece alguna.

Williamson (2000) ha defendido la idea de que el conocimiento no es analizable en términos de creencia. Ello deja abierta la posibilidad de que, de hecho, conocimiento no implique creencia. Curiosamente, como Williamson mismo discute, una fuente de resistencia a la concepción del conocimiento como un tipo de estado mental primitivo se basa en la idea que tenemos acceso privilegiado a nuestros estados mentales actuales. De acuerdo con esta idea, "uno debe estar en posición de saber si está en un estado mental, al menos cuando está atendiendo a la cuestión" (p. 93), esto es, que nuestros estados mentales, o al menos algunos de ellos, son luminosos. Pero si el conocimiento es un estado mental y nuestros estados mentales son luminosos, entonces, ello nos comprometería con el principio KK, según el cual un agente sabe que sabe lo que sabe, principio contra el que el propio Williamson (2000) ha argumentado contundentemente.

Para mostrar que no hay tales estados luminosos, Williamson (sección 4.3, p. 96) considera un estado en el cual alguien siente frío y que podemos adaptar, *mutatis mutandi* para los propósitos que nos conciernen, al caso de una creencia de primera persona como la creencia de que yo tengo frío. Podemos imaginar que la temperatura va aumentando muy lentamente de forma que, varias horas después, creo que tengo calor. Uno pasa de creer que uno mismo tiene frío a creer que uno mismo no tiene frío, y de estar en una posición de saber que uno mismo cree que uno mismo tiene frío a estar en una posición de saber que uno mismo cree que uno mismo no tiene frío.

Supongamos que la sensación de calor y frío cambia tan despacio que uno no se percata de ningún cambio en el transcurso de un milisegundo. Supongamos, además, que el sujeto está constantemente considerando su sensación de frío o calor y, por tanto, formando la correspondiente creencia, de manera que respondería a la pregunta "¿tienes frío?" primero con un rotundo "sí", horas después con vacilación, y finalmente con un rotundo "no". Bajo la plausible asunción de que si uno sabe que tiene frío en t, entonces, después de una fracción de tiempo (δ) tan pequeña como se quiera se sigue teniendo frío, Williamson construye un argumento sorítico para mostrar que no hay estados luminosos. Si el estado es luminoso y el sujeto considera a profundidad si lo tiene, entonces, si en $t+\delta$ el sujeto tiene frío, en $t+\delta$ el sujeto sabe que tiene frío. Pero iterando el razonamiento llegamos a la contradicción de que horas después de t, el sujeto tiene frío y no tiene frío. Si Williamson está en lo correcto, el razonamiento de Vidal no puede estarlo puesto que se compromete con la existencia de estados hiperluminosos y, como hemos visto, la hiperluminosidad es una condición aún más fuerte que la luminosidad y que la implica.

Es importante señalar aquí que el argumento de Williamson en contra de la existencia de estados luminosos es controvertido.[10] No obstante, las consideraciones acerca de la luminosidad son interesantes pues Vidal se compromete a algo muchísimo más fuerte. El breve análisis expuesto aquí no pretende mostrar que el argumento de Vidal no sea sólido, ni que su noción de conocimiento no sea coherente, sino meramente exponer, por un lado, la relevancia y singularidad de su resultado, y por otro, las dificultades que enfrenta la noción de conocimiento que Vidal parece tener y sobre la cual descansa el éxito de su argumento. Discutiré a continuación los detalles del argumento.

3.2. Conocimiento y creencia: revisión del argumento

CONOCERME muestra que las creencias de primera persona son *hiperluminosas*. Este resultado está en tensión con el argumento de Williamson en contra de la luminosidad de los estados mentales. Sin embargo, varios autores han presentado objeciones al argumento de Williamson, por lo que en esta sección me centraré en analizar el argumento de Vidal. Mostraré que el paso que resulta inválido es inferir del hecho de que X crea que ella misma Φ, que X sepa que la creencia de que alguien Φ es acerca de quien cree que alguien Φ, esto es, ella misma. La justificación de esta inferencia, recordemos, nos la da la IER. Como señalaba Rosenthal, en el ejemplo del espejo

[10] Véase, por ejemplo, Berker 2008.

"no puedo estar equivocado [en cuanto a] de quién yo considero que es el reflejo; considero que el reflejo es del mismo individuo que está haciendo esa consideración". Si llamamos '*p*' a la proposición *que la creencia de que alguien* Φ *es acerca de quien cree que alguien* Φ, *es decir, ella misma*, Vidal estaría afirmando que "*X* no puede ignorar *p*" (p. 41). Esta aseveración admite dos posibles lecturas, una débil, según la cual *X* no puede formarse la creencia de que no *p*, que estaría justificada pero que no da soporte a la inferencia, y una fuerte, según la cual *X* de hecho siempre se forma la creencia de que *p*, que da soporte a la inferencia, pero no está justificada.

La lectura débil está soportada por ejemplos como el de Rosenthal. *X* no puede equivocarse acerca de la identidad en cuestión, pero eso no es suficiente para tener conocimiento si éste requiere creencia. Parece plausible pensar que si *S* se forma la creencia de que *p*, entonces, esta creencia es siempre verdadera, está justificada y formada adecuadamente. Pero eso no basta para deducir 3 de 2 (recordemos: para deducir de que (2) *X* cree que ella misma Φ que (3) *X* sabe que la creencia de que alguien Φ es acerca de quien cree que alguien Φ, es decir, ella misma), pues el sujeto puede no formarse la creencia. Si conocimiento implica creencia, entonces, aun cuando aceptemos la imposibilidad de error y la imposibilidad de formarse la creencia de forma errónea, es algo más lo que se requiere para que haya conocimiento de que α; hace falta, además, formarse la creencia de que α y el sujeto puede perfectamente no hacerlo.

Por el contrario, alguien podría defender una lectura fuerte de '*X* no puede ignorar *p*' según la cual de hecho no la ignora y sabe que *p*, afirmando que el sujeto siempre se forma esa creencia. Me parece, no obstante, que tan sólo con que aceptemos que la formación de una nueva creencia tiene una demanda cognitiva no nula y, por tanto, que no podemos multiplicar libremente las creencias que supuestamente tenemos, tal posición resulta implausible e injustificada. Más aún, esta consideración tiene consecuencias inaceptables cuando lo combinamos con SC*. CONOCERME establece que si *X* cree que α, entonces *X* sabe que ella misma cree que α. Si conocer implica creer, entonces, *X* cree que ella misma cree que α y de nuevo por CONOCERME *X* sabe que ella misma cree que ella misma cree que α y así *ad infinitum*.[11,12] Cuando combinamos este conjunto infinito de

[11] Agradezco a Moisés Vaca por llamar mi atención a este hecho.

[12] En su nota 8, Vidal anticipa un problema similar de su propuesta y señala, en respuesta, que cree que los estados intencionales se individuan en términos de su rol funcional y que "es muy posible que a partir del conocimiento de segundo

creencias con SC* vemos que todas ellas satisfacen el antecedente y, por tanto, de ser SC* verdadero, todas ellas son creencias conscientes. Ahora bien, no veo ningún sentido relevante de "consciente" bajo el cual resulte plausible que cuando creo conscientemente que voy a ver a mi abuela, tengo una infinidad de creencias conscientes del tipo "creo que yo mismo creo que yo mismo creo (. . . y así infinitamente. . .) que yo mismo creo que yo mismo voy a ver a mi abuela".

Alternativamente uno puede negar el vínculo entre conocimiento y creencia. Este paso hace que el resultado de Vidal resulte mucho menos interesante, pues la gran mayoría de los epistemólogos estarían de acuerdo con que conocimiento implica creencia y, como hemos visto, basta esta implicación para bloquear su argumento.[13] En cualquier caso, a continuación proseguiré con la discusión del argumento bajo el supuesto de que saber no implica creer.

Williamson defiende la posición que permite romper el vínculo entre creencia y conocimiento pero, como hemos visto, también rechaza la luminosidad de tales estados y *a fortiori* la hiperluminosidad. Quizá alguien encuentre atractiva la propuesta de Williamson, pero no sus razones para negar la luminosidad, o tal vez podamos desligar conocimiento de creencia si mantenemos que alguna forma de justificación proposicional (Turri 2010) bastará para tener conocimiento. En este caso, el sujeto sabe todas aquellas proposiciones tales que si las creyera, la creencia estaría bien formada y justificada. De este modo, aun cuando X no se forme la creencia de que p, dado que en caso de formársela ésta estaría bien formada y estaría justificada, entonces X sabe que p. Bajo estas circunstancias, CONOCERME parece un argumento sólido, aunque traslada el problema a la segunda parte

orden no estaríamos describiendo ya ningún aspecto determinante del rol funcional de ese estado". Adaptando esa idea una posible forma de bloquear este argumento sería mantener que en realidad la creencia de tercer orden (y todas las demás) y la de segundo orden son en realidad un mismo estado. Sin embargo, parece que hace falta, cuando menos, algún tipo de argumento o razón que justifique que de hecho no juegan roles funcionales distintos y Vidal no ofrece ni sugiere ninguno.

[13] El argumento de Vidal, como él mismo observa (2014, nota 2), puede ser reformulado *mutatis mutandi* para llegar a la conclusión de que "todos los estados intencionales con contenido de primera persona, los pensamientos o actitudes *de se* (Lewis 1979), son necesariamente conscientes". Como un árbitro me ha señalado, si el conocimiento es un estado intencional, entonces no importa si está vinculado con una creencia. Según CONOCERME, el conocimiento que yo mismo creo que X, al ser un estado intencional, requeriría un conocimiento de que yo mismo sé que creo que X, y así *ad infinitum*. No quiero presionar más allá de este punto pues no tengo claros los compromisos cognitivos de una teoría del conocimiento que no implique creencia y, por tanto, si es coherente o no pensar que podríamos tener tal conocimiento.

de su argumento: el problema en este caso es justificar la conexión entre conocimiento y consciencia, esto es, entre el conocimiento que *X* tiene de la creencia de primera persona y el hecho de que la creencia esté disponible para *X*.

Para poder entender la justificación de esa conexión, debemos hacer explícita la noción de consciencia relevante pues, como es sabido, resulta ambigua. La noción de consciencia relevante será aquella determinada por el debate en que toman parte las teorías de orden superior. De otra forma, aunque el argumento de Vidal fuera sólido y mostrara que las creencias de primera persona son, en algún sentido, estados conscientes, si tal noción no refiriera a la propiedad de la que hablan las teorías de orden superior y fuera distinto del sentido en el que requieren que haya pensamientos de primera persona que no son conscientes, el defensor de las teorías de orden superior no tendría nada de que preocuparse.

Ned Block (1995) introdujo una distinción conceptual entre dos nociones de consciencia. Por un lado, tenemos lo que Block llama 'consciencia fenoménica': un estado es fenoménicamente-consciente cuando hay algo que es para el sujeto estar en ese estado. Por otro, un estado es 'acceso-consciente', *grosso modo*, cuando su contenido está disponible para la formación de creencias y el control racional de la acción. Una crítica común a esta distinción es que no parece que ambas nociones puedan ser confundidas, pues la primera es una propiedad ocurrente y la segunda meramente disposicional.[14] Por esta razón y para que sea una noción funcional susceptible de ser identificada con la de consciencia fenoménica, Block (2007, 2011b) ha afinado la noción de acceso-consciente explicando la "disponibilidad" (propiedad disposicional) en términos de "difusión" (*broadcast*, propiedad categórica). Una vez afinada de este modo, es una pregunta abierta si consciencia fenoménica y consciencia de acceso refieren a propiedades distintas.[15] Independientemente de la respuesta a esta cuestión —que los defensores de las teorías de orden superior tienden a responder de forma negativa (Rosenthal 2007)—, Vidal pretende rebatir las teorías de orden superior y estas teorías —explícitamente la de Rosenthal, que es el objetivo principal de las críticas de Vidal— son teorías "*ambiciosas*" de la consciencia (Block 2011a), es decir, teorías que pretenden explicar la consciencia fenoménica (Rosenthal 2011; Weisberg 2011).

[14] Para una elaboración más detallada de este punto, véanse Burge 1997; Kriegel 2009.

[15] Para un debate más profundo, véanse Block 2007 y 2011b; Philips 2011; Stazicker 2011, y Sebastián 2014.

La noción de consciencia a la que se refiere Vidal es cercana a la de acceso —lo cual no es en sí problemático si uno cree que consciencia de acceso y fenoménica capturan la misma propiedad, cuestión en la que no ahondaré aquí—. Citando a Shoemaker (2009), Vidal señala que "una creencia es consciente, en un sentido psicológicamente (o funcionalmente) relevante, cuando está *disponible* para guiar el razonamiento y la acción" (nota 17, p. 48). Ahora bien, en sintonía también con Shoemaker, enfatiza igualmente que tal disposición es independiente de que la creencia sea o no ocurrente (algo de lo que hace uso en su argumentación). El problema es que eso desvía la noción de Shoemaker de cualquier noción relevante en el debate en que toman parte las teorías de orden superior, ya que éste requiere que el estado sea ocurrente.[16] En contra de lo que afirma Shoemaker, no tengo en absoluto claro en qué sentido un estado meramente disposicional puede contar como estado consciente. Por suerte, mi claridad al respecto es irrelevante en el argumento y lo importante es la propiedad que las teorías de orden superior pretenden explicar, a saber, la consciencia fenoménica. Debido a que los estados fenoménicamente conscientes son estados ocurrentes, la noción de consciencia de Shoemaker no parece cumplir con este requisito y, por tanto, no ser relevante para la discusión.[17] Podemos, en consecuencia, dejar a un lado por el momento estados disposicionales y considerar que la noción de consciencia en juego es algo semejante a la noción de acceso-consciencia que los defensores de las teorías de orden superior ambiciosas, como Rosenthal, consideran que va de la mano de la de consciencia fenoménica (Brown y Lau en prensa; Rosenthal 2007).

Una vez aclarada la noción de consciencia relevante, podemos pasar a analizar el vínculo entre conocimiento y consciencia. Vidal propone que tal vínculo esté dado por SC*, principio que, recordemos, afirma que cuando una creencia cumple cierta condición (que si X cree que α, entonces X sabe que ella misma cree que α), entonces, esa creencia ha de ser consciente. Las creencias de primera persona cumplen la condición dada por el antecedente si CONOCERME es un argumento sólido. Sin embargo, lo que falta para dar soporte a SC* es encontrar una razón que justifique que una creencia que cumple la condición expresada por el antecedente tenga que ser una creencia consciente.

[16] Incluso en el caso de la consciencia de acceso, aunque no se acceda de hecho al contenido del estado, sí se ha de difundir. Esto ocurre en virtud de un estado en que el sujeto de hecho está.

[17] Rosenthal (2005) hace explícito este punto en su crítica a las teorías de orden superior disposicionales. Para una presentación concisa y breve de esta discusión, véase Carruthers 2016.

Esto es relevante porque la noción de creencia de primera persona inconsciente —o en general de actitud proposicional de primera persona— no parece en absoluto incoherente y durante años se ha hablado de creencias de primera persona inconscientes. Por ejemplo, Freud planteó que los hombres en algún momento de sus vidas tienen el deseo de primera persona inconsciente de cohabitar con sus madres y el deseo de primera persona inconsciente de matar a sus padres. Pero, al menos *prima facie*, no parece que esas ideas resulten incoherentes por el hecho de involucrar actitudes proposicionales de primera persona —nótese que CONOCERME se puede reproducir, *mutatis mutandi*, para cualquier actitud proposicional de primera persona, no sólo creencias, como el propio Vidal (2014) menciona explícitamente—. Pero tales creencias son, por definición, inconscientes en contra de lo que mantiene SC*. Obviamente, si SC* es verdadero, entonces, tales creencias no existen, pero la discusión requiere que SC* pueda ser motivada independientemente. Sin embargo, todo lo que Vidal (2014) nos dice es lo siguiente:

> [S]i el antecedente del condicional (SC*) es verdadero para cierto tipo de creencia, eso significa que la posesión de una creencia de tal tipo no puede ocurrir sin la posesión del conocimiento de que uno mismo tiene esa creencia. Diríamos también que las creencias de tal tipo no tienen un rol funcional en la psicología de una persona que sea independiente de ese conocimiento. Entonces, ese conocimiento no puede ser el de una creencia que, sin embargo, es inconsciente: la creencia es consciente, en el sentido de creer conscientemente algo, porque no tiene una realidad constituida psicológicamente al margen de ese conocimiento. (p. 46)

Esta aseveración parece totalmente insuficiente, pues no hay conexión conceptual entre el hecho de que el rol funcional en la psicología de una persona dependa del conocimiento que uno tiene y el hecho de que ese estado esté disponible en el sentido relevante para *X*. Parece que no hay tensión conceptual en la idea de tener conocimiento inconsciente. Parece plausible sostener que yo tengo en todo momento la creencia de que vivo en la Ciudad de México, de que trabajo en la UNAM, de que creo que dos más dos son cuatro y que de hecho sé en todo momento que creo que yo vivo en la Ciudad de México, que creo que trabajo en la UNAM, o que creo que dos más dos son cuatro. Ahora bien, no hay un sentido relevante en que se pueda afirmar que yo tenga en todo momento esas creencias de forma consciente. Quizá haya un sentido en el que la creencia esté en todo momento disponible para la formación de otras creencias y el control de la

acción, pero no parece distinto del sentido en que lo pueda ser mi creencia inconsciente de que, digamos, París es la capital de Francia. Por ejemplo, las creencias anteriores plausiblemente permiten que me forme la creencia de que en la ciudad en que vivo llueve mucho en verano (dada mi creencia de que vivo en la Ciudad de México y mi creencia de que llueve mucho en la Ciudad de México), o para decidir salir a pasear después del trabajo (dada mi creencia de que trabajo en la UNAM y mi creencia de que en la UNAM hay un parque). En contra de lo que creo, tal vez haya un uso de "consciencia" que sea legítimo en virtud de tal disposición, pero entonces la tesis de que las creencias de primera persona son conscientes, en tal sentido, no parece demasiado interesante y, desde luego, no valdría para poner en jaque las teorías de orden superior, pues se trata claramente de otra noción de consciencia, distinta tanto de la de acceso-consciencia como de la de consciencia fenoménica. Antes de escribir estos párrafos, ni mi creencia de que París es la capital de Francia ni la de que trabajo en YY habían sido "difundidas", digamos en la memoria de trabajo (Block 2007; Sebastián 2014), para la formación de otras creencias, ni había nada que era para mi tener esas creencias.[18]

Para que el resultado sea interesante para el debate, debemos mostrar una conexión entre consciencia y conocimiento. Sin embargo, la conexión al nivel cuasiintuitivo al que apela Vidal sólo parece mantenerse si la noción de consciencia es una que no es relevante. Cabe entonces buscar un argumento a favor de SC* donde la noción de consciencia sea la que nos interesa.

Un posible argumento sería el siguiente (tengo la impresión de que algo parecido está detrás del razonamiento de Vidal):

i) X cree que $\alpha \rightarrow$ X sabe que ella misma cree que α (CONOCERME)

ii) X cree que α (asunción)

iii) X sabe que ella misma cree que α (de i y ii)

iv) X cree que ella misma cree que α (saber implica creer)

v) 'X cree que α' está disponible para X

vi) X cree que α conscientemente (de v) y definición de consciente)

[18] Seguramente porque tales creencias eran meros estados disposicionales (Schwitzgebel 2014) y, por lo tanto, estados no susceptibles de ser ni acceso, ni fenoménicamente conscientes.

> (X cree que $\alpha \rightarrow$ X sabe que ella misma cree que α) $\rightarrow$ (X cree que $\alpha \rightarrow$ X cree que α conscientemente)

Sin embargo, este argumento no es sólido. En primer lugar, no está claro cómo conectar el hecho de que si X cree que ella misma cree que α entonces la creencia esta disponible para X, pero aún cuando aceptemos ese paso, el argumento presenta un problema mayor —un razonamiento análogo al que hace Vidal— que hace que resulte inválido: el paso de iii) a iv) requiere que saber implique creer y como hemos visto CONOCERME no es válido bajo esa suposición. Tal vez, entonces un principio suficientemente cercano fuera aceptable:

> SC**: Si (si X cree que α, entonces X cree que ella misma cree que α) entonces (si X cree que α, entonces X cree que α conscientemente)
>
> En este caso:
>
> i′) X cree que $\alpha \rightarrow X$ cree que ella misma cree que α (asunción)
>
> ii′) X cree que α (asunción)
>
> iii′) X cree que ella misma cree que α (de i′ y ii′)
>
> iv′) 'X cree que α' está disponible para X
>
> v′) X cree que α conscientemente (de iv′) y definición de consciente)
>
> ———————
>
> (X cree que $\alpha \rightarrow$ X cree que ella misma cree que α) $\rightarrow$ (X cree que $\alpha \rightarrow$ X cree que α conscientemente)

Efectivamente, en este caso si aceptamos el paso de iii′ a iv′, el argumento parece sólido. El problema es que no tenemos justificación para creer que las creencias de primera persona satisfacen la premisa i′), donde α es una creencia de primera persona. Como hemos visto, IER no le da soporte, pues de la imposibilidad de la ignorancia no se sigue que el sujeto de hecho se forme la creencia adecuada, sólo que no se puede formar la errónea. Más aún, el defensor de la teoría de orden superior acepta felizmente que los casos en los que el sujeto de hecho se forma la creencia de orden superior (X cree que ella misma cree que α) y que por tanto no hacen el antecedente falso, son casos en los que el consecuente es verdadero y que de hecho,

en esas circunstancias, el sujeto tiene una creencia consciente: eso es precisamente lo que la teoría predice.

No pretendo sostener que este argumento es la única forma de mostrar que SC*, o un principio suficientemente cercano, sea verdadero, pero sí que muestra un problema. CONOCERME me permite concluir, en el caso de las creencias de primera persona, que si las creo, entonces sé que las creo y éste es el antecedente del principio que queremos demostrar. Para justificar el consecuente, he de conectar el hecho de que conozca la creencia con el hecho de que esté disponible para *X* en el sentido relevante (que es de lo que la consciencia depende). Si lo hago por medio de la creencia correspondiente, estoy asumiendo que conocer implica creer, como hago en el argumento. Si por otro lado, voy directamente de iii) a v) entonces estaré asumiendo directamente aquello que quiero mostrar (a saber, que hay una conexión entre conocimiento y consciencia) y, por lo tanto, prejuzgando la cuestión en contra de aquel que cree que SC* no es verdadero.

4. *Conclusión*

Vidal ha argumentado que necesariamente toda creencia de primera persona ha de ser una creencia consciente. En este artículo he mostrado que el argumento es inválido si la noción de consciencia en juego es una noción relevante, esto es, una noción susceptible de poner en jaque las teorías de orden superior o aquellas teorías que explican ciertos aspectos del comportamiento postulando estados mentales inconscientes.

En particular, he mostrado que el argumento que permite concluir que conocemos toda creencia de primera persona que tenemos es inválido bajo la tesis ampliamente aceptada de que conocer implica creer. Obviamente mi oponente no tiene por qué aceptar este principio, pero en ausencia de un argumento que muestre que conocer no implica creer, el alcance de su argumento resulta muy restringido, pues, como he mostrado, basta aceptar que conocer implica creer para bloquear su argumento. La noción de conocimiento que surge resulta cuando menos digna de sospecha ya que toma el conocimiento como primitivo y requiere algo aún más fuerte que la luminosidad.

Finalmente, incluso si esta noción de conocimiento es consistente o hasta verdadera, muestro que no hay ninguna razón para vincular tal conocimiento y consciencia, como pretende Vidal, al menos si la consciencia ha de ser entendida de forma que la conclusión del argu-

mento sea significativa para las tesis que mantienen los defensores de la teorías de orden superior.[19]

BIBLIOGRAFÍA

Berker, S., 2008, "Luminosity Regained", *Philosophers' Imprint*, vol. 8, no. 2, pp. 1–22.

Block, N., 2011a, "The Higher Order Approach to Consciousness Is Defunct", *Analysis*, vol. 71, no. 3, pp. 419–431.

———, 2011b, "Perceptual Consciousness Overflows Cognitive Access", *Trends in Cognitive Sciences*, vol. 15, no. 12, pp. 557–575.

———, 2007, "Consciousness, Accessibility and the Mesh between Psychology and Neuroscience", *Behavioral and Brain Sciences*, vol. 30, no. 4, pp. 481–548.

———, 2002, "On a Confusion about the Function of Consciousness", en N. Block (comp.), *Consciousness, Function and Representation: Collected Papers*, vol. 1, Bradford Books, Bradford. (Primera edición 1995, primera reimpresión 2002.)

Block, N., O. Flanagan, y G. Guzeldere (comps.), 1997, *The Nature of Consciousness: Philosophical Debates*, MIT Press, Cambridge, Mass.

Brown, R., 2011, "The Myth of Phenomenological Overflow", *Consciousness and Cognition*, vol. 21, no. 2, pp. 599–604.

Brown, R. y H. Lau, en prensa, "The Emperor's New Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations", en A. Pautz y D. Stoljar, (comps.), *Festschrift for Ned Block*, MIT Press, Cambridge, Mass.

Burge, T., 1997, "Two Kinds of Consciousness", en Block, Flanagan y Guzeldere 1997, pp. 427–435.

Cappelen, H. y J. Dever, 2013, *The Inessential Indexical. On the Philosophical Insignificance of Perspective and the First Person*, Oxford University Press, Nueva York.

Carruthers, P., 2016, "Higher-Order Theories of Consciousness", *The Stanford Encyclopedia of Philosophy* (Fall 2016 Edition), Edward N. Zalta (ed.), disponible en: <https://plato.stanford.edu/archives/fall2016/entries/consciousness-higher/> [última consulta: 17/03/2017].

Castañeda, H.N., 1966, " 'He': A Study in the Logic of Self-Consciousness", *Ratio*, no. 8, pp. 130–157.

[19] Estoy enormemente agradecido con Pedro Stepanenko y Francisco Vidal por la discusión que originó e hizo posible este trabajo. También quiero agradecer profundamente a Miguel Ángel Fernández, a Angélica Pena-Martínez y a los dos árbitros anónimos sus comentarios sobre un borrador anterior. Este trabajo fue presentado dentro del Seminario de Investigadores del Instituto de Investigaciones Filosóficas de la UNAM, agradezco a los participantes por sus comentarios y sugerencias, especialmente a Axel Barceló, Maite Ezcurdia, Eduardo García Ramírez y Moisés Vaca.

Chierchia, G., 1989, "Anaphora and Attitudes *De Se*", en R. Bartsch, J. van Benthem y P. van Emde Boas (comps.), *Semantics and Contextual Expression*, Foris, Dordrecht, pp. 1–31.

Chisholm, R.M., 1981, *The First Person: an Essay in Reference and Intentionality*, University of Minnesota Press, Minneapolis.

Corazza, E., 2004, *Reflecting the Mind: Indexicality and Quasi-Indexicality*, Clarendon Press, Oxford.

Dowty, D., 1985, "On Recent Analyses of the Semantics of Control", *Linguistics and Philosophy*, vol. 8, no. 2 pp. 291–331.

Ezcurdia, M., 2001, "Thinking about Myself", en A. Brook (comp.), S*elf-Reference and Self-Awareness*, John Benjamins, Amsterdam, pp. 179–203.

Feit, N., 2008, *Belief about the Self: A Defense of the Property Theory of Content*, Oxford University Press, Nueva York.

Finkelstein, D., 2003, *Expression and the Inner*, Harvard University Press, Cambridge, Mass.

Gennaro, R., 2012, *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*, MIT Press, Cambridge, Mass.

———, 1996, *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*, John Benjamins, Amsterdam.

Kaplan, D., 1989, "Demonstratives", en J.P.J. Almog y H. Wettstein (comps.), *Themes from Kaplan*, Oxford University Press, Nueva York, pp. 481–563.

King, J., S. Soames y J. Speaks, 2014, *New Thinking about Propositions*, Oxford University Press, Nueva York.

Kriegel, U., 2009, *Subjective Consciousness: A Self-Representational Theory*, Oxford University Press, Nueva York.

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

O'Brien, L., 2007, *Self-Knowing Agents*, Oxford University Press, Oxford.

Perry, J., 1980, "Belief and Acceptance", *Midwest Studies in Philosophy*, vol. 5, no. 1, pp. 533–542.

———, 1979, "The Problem of the Essential Indexical", *Noûs*, vol. 13, no. 1, pp. 3–21.

Phillips, I.B., 2011 "Perception and Iconic Memory: What Sperling Does Not Show", *Mind and Language*, vol. 26, no. 4, pp. 381–411.

Prinz, J., 2012, *The Conscious Brain*, Oxford University Press, Nueva York.

Richard, M., 1983, "Direct Reference and Ascription of Belief", *Journal of Philosophical Logic*, vol. 12, no. 4, pp. 425–452.

Rosenthal, D.M., 2011, "Exaggerated Reports. Reply to Block", *Analysis*, vol. 71, no. 3, pp. 431–437.

———, 2007, "Phenomenological Overflow and Cognitive Access", *Behavioral and Brain Sciences*, vol. 30, no. 4, pp. 521–522.

Rosenthal, D.M., 2005, *Consciousness and Mind*, Oxford University Press, Nueva York.
———, 2004, "Being Conscious of Ourselves", *The Monist*, vol. 87, no. 2, pp. 159–181.
———, 1997, "A Theory of Consciousness", en Block, Flanagan y Guzeldere 1997, pp. 729–755.
Rovane, C., 1987, "The Epistemology of First-Person Reference", *The Journal of Philosophy*, vol. 84, no. 3, pp. 147–167.
Salmon, N., 1989, "Illogical Belief", *Philosophical Perspectives*, vol. 3, pp. 243–285.
Schwitzgebel, E., 2014, "Belief", *The Stanford Encyclopedia of Philosophy*, disponible en <http://plato.stanford.edu/archives/spr2014/entries/belief/> [última consulta: 17/05/2017].
Sebastián, M.A., 2014, "Dreams: An Empirical Way to Settle the Discussion between Cognitive and Non-Cognitive Theories of Consciousness", *Synthese*, vol. 91, no. 2, pp. 263–285.
———, 2012a, "Experiential Awareness: Do You Prefer 'It' to 'Me'"? , *Philosophical Topics*, vol. 40, no. 2, pp. 155–177.
———, 2012b, "Review of U. Kriegel 'Subjective Consciosness: A Self-Representational Theory'", *Disputatio*, vol. 6, pp. 413–417.
Shoemaker, S., 2009, "Self-Intimation and Second Order Belief", *Erkenntnis*, vol. 71, no. 1, pp. 35–51.
———, 1968, "Self-Reference and Self-Awareness", *The Journal of Philosophy*, vol. 65, no. 19, pp. 555–567.
Stalnaker, S., 1999, *Context and Content: Essays on Intentionality in Speech and Thought*, Oxford University Press, Nueva York.
Stazicker, J., 2011, "Attention, Visual Consciousness and Indeterminacy", *Mind and Language*, vol. 26, no. 2, pp. 156–184.
Turri, J., 2010, "On the Relationship between Propositional and Doxastic Justification", *Philosophy and Phenomenological Research*, vol. 80, no. 2, pp. 312–326.
Vidal, J., 2015, "Pensamientos *de se* y mente consciente", *Theoria*, vol. 30, no. 1, pp. 117–136.
———, 2014, "Creencia de primera persona, conciencia y la paradoja de Eroom", *Crítica. Revista Hispanoamericana de Filosofía*, vol. 46, no. 138, pp. 37–64.
Weisberg, J., 2011, "Abusing the Notion of What-It-Is-Like-ness: A Response to Block", *Analysis*, vol. 71, no. 3, pp. 438–443.
Williamson, T., 2000, *Knowledge and its Limits*, Oxford University Press, Nueva York.
Wittgenstein, L., 1958, *The Blue and Brown Books*, Harper and Row, Nueva York.

Recibido el 8 de marzo de 2016; revisado el 7 de marzo de 2017; aceptado el 29 de mayo de 2017.