CRÍTICA, Revista Hispanoamericana de Filosofía. Vol. 46, No. 138 (diciembre 2014): 37–64

CREENCIA DE PRIMERA PERSONA, CONCIENCIA Y LA PARADOJA DE EROOM

JAVIER VIDAL
Departamento de Filosofía
Universidad de Concepción
fravida@udec.cl

RESUMEN: En este artículo se trata de mostrar que existe una relación necesaria, no contingente, entre tener una creencia de primera persona y creer conscientemente: una creencia de primera persona es necesariamente consciente. Pero de aquí se siguen al menos dos importantes consecuencias. Primero, tendría que ser abandonada una teoría de la conciencia según la cual un estado mental es consciente cuando va acompañado por un pensamiento o una creencia de orden superior acerca de ese estado. Segundo, podría darse cierta explicación de la paradoja de Eroom, que es el sinsentido de aseverar o de creer algo de la forma "*p* y yo creo inconscientemente que *p*".

PALABRAS CLAVE: referencia de primera persona, cuasi-indéxico, pensamiento de orden superior, creencia inconsciente, unidad de la conciencia

SUMMARY: This paper aims to show that there exists a necessary, non-contingent, relation between having a first-person belief and believing consciously: a first-person belief is necessarily conscious. From this, two major consequences can be drawn. First, a theory of consciousness claiming that a mental state is conscious when it is accompanied by a higher-order thought or belief about the state itself should be discarded. Second, an account can be given of Eroom's paradox —the nonsense of asserting or believing something of the form "*p* and I believe unconsciously that *p*"—.

KEY WORDS: first-person reference, quasi-indexical, higher-order thought, unconscious belief, unity of consciousness

I

Mi objetivo es mostrar que existe una relación necesaria, no contingente, entre tener una creencia de primera persona y creer conscientemente: una creencia de primera persona es necesariamente, a causa de tener un contenido de primera persona de la forma "Yo Φ",[1] una creencia consciente.[2] Es decir, una creencia de primera persona

[1] La variable predicativa "Φ" está por cualquier verbo, conjugado en tiempo pasado, presente o futuro, usado para referirse a estados, acciones y, en general, a propiedades físicas o mentales de una persona.

[2] Como veremos, esto significa introducir cierto sentido, psicológicamente (o funcionalmente) relevante, de "creer conscientemente". En realidad, considero que el tipo de argumento que voy a plantear a favor de la naturaleza consciente, en ese

no puede ser inconsciente precisamente porque el análogo mental del pronombre personal "yo" es instanciado en esa creencia. En la segunda parte argumentaré, a partir de la tesis de que no puede haber creencias de primera persona que sean inconscientes, contra cierto tipo de teoría de orden superior de la conciencia (Higher-Order Thought Theory of Consciousness) según la cual la naturaleza consciente de un estado mental consiste en tener un pensamiento o una creencia de orden superior acerca de ese estado. Finalmente, en la tercera parte presentaré, a partir de esa misma tesis, una explicación de la paradoja de Eroom, que es el sinsentido de aseverar o de creer algo con un contenido de la forma "*p* y yo creo inconscientemente que *p*".

Consideremos una creencia de primera persona, es decir, una creencia con un contenido de primera persona de la forma:

(1) Yo Φ.

Entonces, la adscripción de una creencia de primera persona es de la forma:

(2) *X* cree que ella misma Φ.

El pronombre reflexivo, o cuasi-indéxico, "ella misma/él mismo" es la traducción a contextos indirectos como (2) del pronombre personal "yo", usado en cualquier sustitución de (1) para expresar directamente el contenido de una creencia de primera persona (Geach 1957).[3] Ahora bien, el cuasi-indéxico tiene la función de eliminar, cuando es posible, la ambigüedad de una adscripción de creencia de la forma:

sentido, de las creencias de primera persona puede reformularse directamente para alcanzar la conclusión de que todos los estados intencionales con un contenido de primera persona, los pensamientos o actitudes *de se* (Lewis 1979), son necesariamente conscientes: las creencias, deseos e intenciones, entre otros estados intencionales, con un contenido de la forma "Yo Φ" son necesariamente conscientes. Pero sólo me interesa presentar el argumento para las creencias de primera persona teniendo en cuenta las consecuencias que quiero examinar en las partes segunda y tercera de este artículo.

[3] Como ha sido señalado en la literatura, no todos los usos de "ella misma/él mismo" son como cuasi-indéxico: por ejemplo, "Su impulsividad con frecuencia rebotó contra él mismo". Este ejemplo muestra que "él mismo" no siempre es utilizado para reportar una actitud *de se*. Además, frecuentemente las adscripciones de creencia de primera persona y otras actitudes *de se* no contienen una cláusula-que con un uso del cuasi-indéxico sino una construcción de infinitivo con un sujeto implícito o sobreentendido, como en "*X* cree estar enfermo". Pero los lingüistas representan estas adscripciones en términos de una referencia anafórica que va desde la posición de sujeto del complemento, PRO, hasta la posición de sujeto de toda la oración de tal modo que, además, PRO tiene la función de indicar que no se trata

(2*) X cree que ella Φ.

Pues, una adscripción como (2*) no entraña que X tenga una creencia de primera persona y, por tanto, es ambigua en ese sentido. En efecto, supongamos que X, quien *no* es la persona retratada en cierta fotografía, tiene la creencia de que la persona retratada en la fotografía es deportista. Entonces, es correcta una adscripción de la forma (2*): constituye una propiedad del pronombre "ella/él" poder reemplazar *salva veritate*, en contextos indirectos como la adscripción de una creencia, cualquier término singular o descripción definida, como "la persona retratada en la fotografía", que pudiera usarse también para expresar directamente el contenido de la creencia. Por ello, puede pasarse de reportar que X cree que la persona retratada en la fotografía es deportista a reportar que X cree que ella es deportista. Sin embargo, X no es la persona retratada en la fotografía. Más aún, supongamos que X *es* la persona retratada en la fotografía pero no se reconoce en ella. Según el principio *salva veritate* aplicado ahora a un uso anafórico del pronombre "ella", también ahora puede pasarse de reportar que X cree que la persona retratada en la fotografía es deportista a reportar que X cree que ella es deportista. Empero, X ignora que ella misma es la persona retratada en la fotografía. En ninguno de los dos casos X tiene una creencia de primera persona y, por ello, no es correcta una adscripción de la forma (2) (Bermúdez 1998, pp. 2–4; Castañeda 1966/1999; Corazza 2004, pp. 279–285).

Parece así que, para dar cuenta de la función del cuasi-indéxico, es necesario suponer que el pronombre personal "yo" es usado en cualquier sustitución de (1) y el análogo mental de "yo" es instanciado en una creencia para referirse a uno mismo *como uno mismo* (Castañeda 1989), lo que significa referirse a uno mismo de tal modo que la referencia de primera persona está garantizada. Pues, obviamente, el referente de "yo" o de su análogo mental en la creencia, el hablante o el creyente, necesariamente existe, en cuyo caso siempre hay un referente, uno mismo, a quien uno siempre se refiere con éxito como uno mismo (Rovane 1987, p. 153). A este respecto, Lucy O'Brien dice:

---

de una simple anáfora sino de la adscripción de una actitud *de se* (Chierchia 1989). De manera que la adscripción anterior sería representada así:

> X cree PRO estar enfermo.

Entonces, el sujeto implícito PRO de las construcciones de infinitivo debe entenderse como un cuasi-indéxico y este hecho es considerado una razón a favor de la existencia de los cuasi-indéxicos en el lenguaje natural (Corazza 2004, pp. 280–281).

> Parece que la referencia de primera persona está garantizada, y garantizada de tres modos. En primer lugar, un sujeto siempre tiene éxito en referir cuando ella refiere en primera persona. En segundo lugar, ella también tiene éxito en referirse a ella misma cuando refiere en la modalidad de primera persona. En tercer lugar, ella sabe que se está refiriendo a ella misma. Parece que la referencia de primera persona siempre tiene un referente, siempre es una referencia reflexiva y siempre es una referencia reflexiva autoconsciente. (2007, p. 5)[4,5]

Pero, dado que la referencia está garantizada en el sentido reflexivo de que uno siempre se refiere con éxito a uno mismo, es imposible que uno no sea idéntico al referente del análogo mental de "yo" en la creencia. Y dado que la referencia está garantizada en el sentido autoconsciente de que uno siempre se refiere con éxito a uno mismo como uno mismo, es imposible que uno ignore la identidad que guarda con el referente del análogo mental de "yo" en la creencia.[6] Precisamente por eso, la verdad de una adscripción como (2) no sólo se caracteriza porque, como en la creencia X se refiere con éxito a ella misma, es imposible que la persona referida por "X" en la adscripción (fuera de la cláusula-que) no sea idéntica a la persona referida por "ella misma" en la adscripción (dentro de la cláusula-que): es imposible que el sujeto de la creencia no sea idéntico al objeto de la creencia. Es que, además, la verdad de (2) se caracteriza

[4] La traducción al castellano de todas las citas es mía.

[5] Ciertamente, éste es un compromiso con una concepción robusta de la referencia de primera persona, que no es unánimemente aceptada, según la cual se trata de una referencia que entraña cierto modo de presentación de uno mismo consistente en pensarse *como uno mismo* (Ezcurdia 2001). Aunque no puedo ocuparme aquí de defender esta concepción, me permito hacer dos comentarios. Primero, se requiere esta concepción, como he argumentado brevemente, para hacerse cargo de una caracterización de la primera persona motivada por el análisis lingüístico de las adscripciones de actitudes *de se* mediante el cuasi-indéxico "ella misma/él mismo". Segundo, esta concepción ha recibido una aceptación suficientemente amplia entre autores, que no voy a enumerar aquí, que tienen ideas muy diversas sobre el modo de presentación de uno mismo como uno mismo (Ezcurdia 2001, p. 179) y, por tanto, no presupone ninguna teoría específica sobre el significado y la referencia del pronombre personal "yo" ni de su análogo mental. No obstante, diré algo más sobre esta cuestión en la nota 9 y, lo que es más relevante, plantearé una discusión con una concepción menos robusta de la referencia de primera persona en la segunda parte.

[6] Edipo es el hijo de Layo, pero Edipo puede tener la creencia de que el hijo de Layo está muerto mientras ignora la identidad que guarda con el hijo de Layo. Pues, aunque en cierto sentido Edipo se estaría refiriendo a él mismo en la creencia, con todo, no se estaría refiriendo a él mismo *como él mismo*, esto es, no se estaría refiriendo a él mismo autoconscientemente.

porque, como en la creencia X se refiere con éxito a ella misma *como ella misma*, es imposible que la persona referida por "X" 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): es imposible que el sujeto de la creencia ignore la identidad que guarda con el objeto de la creencia. En otras palabras, de (2) se sigue:

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

Debe tenerse en cuenta que (3) no representa la inmunidad a un error de identificación del uso de "yo" en (1) o de su análogo mental en una creencia de primera persona. En efecto, digamos que el análogo mental de "yo" en la creencia es inmune a un error de identificación cuando es imposible que X, quien cree que ella misma Φ, esté equivocada sobre la identidad de la persona que Φ. De manera que el análogo mental de "yo" no es inmune a un error de identificación cuando es posible que X crea erróneamente que ella misma Φ porque y sólo porque la persona que Φ no es la persona a la que el uso del análogo mental de "yo" refiere, o sea, no es ella misma (Shoemaker 1968/1994, pp. 81–82).[7] Es compatible que, en ese caso, (3) sea verdadero y que, con todo, no haya inmunidad a un error de identificación: X no puede estar equivocada o dejar de tener conocimiento sobre la identidad de la persona *de quien cree* que Φ, que es ella misma, pero está o puede estar equivocada sobre la identidad de la persona que Φ, que en ese caso no es ella misma sino algún otro. Como señala David 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 aún aquí no puedo estar equivocado acerca 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).

[7] Cabe establecer, como hace Shoemaker, una distinción entre inmunidad absoluta e inmunidad circunstancial que no voy a considerar aquí. Tampoco estoy teniendo en cuenta una distinción entre inmunidad a la falsedad e inmunidad a la falta de conocimiento de que uno mismo es quien Φ (O'Brien 2007, pp. 210–214). Dejando de lado estas y otras distinciones, un ejemplo de inmunidad a un error de identificación del análogo mental de "yo" en la creencia de que uno mismo está levantando el brazo es cuando esa creencia es resultado de un conocimiento no observacional de la acción. Por el contrario, si esa creencia está basada en la observación, es posible que X, tomando el brazo observado de otra persona como suyo, crea erróneamente que ella misma está levantando el brazo.

Ahora bien, como según (3) *X* sabe que ella misma es tanto el objeto como el sujeto de la creencia de que alguien Φ, resulta que el cuasi-indéxico "ella misma" puede ocupar la posición tanto del sujeto de la creencia (fuera de la cláusula-que subordinada) como de su objeto (dentro de la cláusula-que subordinada) en la siguiente adscripción de ese conocimiento:

(4) *X* sabe que ella misma cree que ella misma Φ.[8]

Esto significa que existe una relación necesaria, no contingente, entre tener una creencia de primera persona y saber que uno mismo tiene esa creencia.[9] Podría pensarse que alcanzado este punto no queda nada más que argumentar a favor de la naturaleza consciente de las

[8] Podría plantearse la objeción de que el cuasi-indéxico "ella misma" (fuera de la cláusula-que subordinada) tiene que analizarse ahora, como en (2), según los lineamientos propuestos, con el resultado de que (2) entrañaría también que *X* sabe que ella misma sabe que ella misma cree que ella misma Φ. Pero, entonces, el cuasi-indéxico introducido nuevamente aquí tiene que analizarse otra vez según los lineamientos propuestos, y esto da lugar a un regreso al infinito. Mi respuesta a la objeción es aceptar la amenaza del regreso, en el siguiente sentido. Aunque no puedo desarrollar este punto, sostengo la tesis de que, en primer lugar, la creencia de primera persona y el conocimiento de esa creencia *no* son estados intencionales numéricamente distintos. Se trata de un solo estado intencional que cae bajo distintas descripciones, es decir, que puede ser descrito como una creencia de primera persona o como el conocimiento de esa creencia, dependiendo del aspecto del rol funcional de ese estado que uno quiera caracterizar. Luego voy a defender que no puede darse una descripción completa del rol funcional de una creencia de primera persona sin mencionar el conocimiento de esa creencia. Pero si un estado intencional se individúa en términos de su rol funcional, resulta que efectivamente la creencia de primera persona y el conocimiento de esa creencia constituyen un solo estado. Del mismo modo, ese estado intencional puede ser descrito según las infinitas maneras que, como plantea la objeción, se siguen del análisis que he propuesto del cuasi-indéxico "ella misma" en (2), a saber: como conocimiento de segundo orden, de tercer orden y así sucesivamente. Sin embargo, es muy posible que a partir del conocimiento de segundo orden no estaríamos describiendo ya ningún aspecto determinante del rol funcional de ese estado. Ya no habría ninguna diferencia relevante entre describir el estado como un conocimiento de segundo orden o como uno de tercer orden. Por tanto, la amenaza de regreso es doblemente inerme. Primero, es inerme porque el análisis propuesto no entraña que una creencia de primera persona es acompañada por un número infinito de estados de conocimiento. Segundo, es inerme porque el análisis propuesto tampoco entraña que el rol funcional de una creencia de primera persona es tal que sólo pueda ser caracterizado por un número infinito de descripciones.

[9] Considero que la consecuencia de (3) y (4) a partir de (2) no presupone ninguna teoría específica sobre el significado y la referencia del pronombre personal "yo", ni de su análogo mental, sino que es un dato primitivo que tiene que ser explicado por cualquier teoría que reconozca la necesidad de un modo de presentación de uno mismo como uno mismo. Pero, a partir de la reciente distinción entre pensamientos

creencias de primera persona, dada cierta relación necesaria, aparentemente obvia, entre tener conocimiento de una creencia y creer conscientemente:

(SC) Si X sabe que ella misma cree que α, entonces X cree conscientemente que α.[10]

Sin embargo, al considerar especialmente los casos de creencia inconsciente (Moran 2001; Finkelstein 2003), parece que un principio de esta forma genérica es falso. En efecto, examinemos el ejemplo de alguien que, después de haber sido psicoanalizada, llega a saber que ella misma cree que comer es obsceno, que no es una creencia con un contenido de primera persona pero puede haber interactuado

de se implícitos y explícitos (Recanati 2007 y 2009), podría argumentarse que la adscripción de una creencia de primera persona que es implícita no es de la forma (2) y que, por tanto, no pueden derivarse (3) y (4). En efecto, un pensamiento *de se* es implícito cuando uno mismo es un constituyente inarticulado del contenido de ese pensamiento en el sentido de que, aunque uno es el parámetro relevante de la circunstancia de evaluación de ese pensamiento, que es verdadero si y sólo si es verdadero de uno mismo, con todo, en el pensamiento no hay una referencia a uno mismo (menos aún hay una referencia de primera persona, a uno mismo como uno mismo). Entonces, se dice que el pensamiento *concierne* a uno mismo sin ser acerca de uno mismo (Recanati 2009, pp. 256–260). Ciertamente, esto significa que la adscripción de una creencia de primera persona que es implícita no sería de la forma (2), que precisamente mediante el cuasi-indéxico adscribe una referencia de primera persona. Ahora bien, me parece que ésta podría ser precisamente una importante razón para dejar de considerar esas creencias como de *primera persona*. Pues, en este artículo estoy trabajando con la idea de que una caracterización de la primera persona debe responder al hecho de que las adscripciones de creencias de primera persona y otros pensamientos o actitudes *de se* son tales que o bien hacen uso del cuasi-indéxico "ella misma/él mismo" o bien deben entenderse en términos de ese cuasi-indéxico (nota 3). Pero, entonces, todas las creencias de primera persona son explícitas. De hecho, la existencia de creencias o pensamientos *de se* implícitos es una cuestión disputada en la literatura (Morgan 2012; García-Carpintero 2013). En cualquier caso, también podría tomarse la argumentación de este artículo sólo como una argumentación sobre las creencias o pensamientos *de se* explícitos. Pues, si, aun concediendo la existencia de pensamientos *de se* implícitos, los pensamientos *de se* explícitos son necesariamente conscientes, se siguen igualmente tanto la objeción a la teoría de orden superior de la conciencia de Rosenthal como la explicación que propongo de la paradoja de Eroom. En efecto, tanto un pensamiento de orden superior como una creencia de Eroom parecen ser creencias o pensamientos *de se* explícitos, y así es como son tratados en la literatura.

[10] Los esquemas (SC) y (SC*), introducidos más adelante, son esquemas donde "α" es una metavariable que puede sustituirse por distintas variables proposicionales y, como se verá, es relevante para el desarrollo del artículo la sustitución en (SC*) por las variables proposicionales "ella misma Φ" y "p y ella misma cree inconscientemente que p".

con otras creencias de primera persona para producir un trastorno alimentario. En este caso, el conocimiento de esa persona está basado en el diagnóstico del psicoanalista, que es una forma de adquisición de conocimiento mediante el testimonio. Pero, obviamente, esa persona no adquiere mediante el testimonio la creencia de que comer es obsceno. La tarea del psicoanalista fue informarle que ella, inconscientemente, ya tenía esa creencia, que, sin embargo, no estaba dispuesta a expresar de manera espontánea mediante una aserción sincera de "Comer es obsceno".

Es posible que, en una fase posterior de integración personal, la creencia de que comer es obsceno llegue a estar disponible a la conciencia y así la persona pueda primero expresarla y después someterla a una revisión esclarecedora. Pero hay que pensar que, como el conocimiento de que ella misma tiene esa creencia fue adquirido mediante el testimonio y no mediante algún tipo de relación directa e inmediata con la creencia, en la fase inicial del descubrimiento psicoanalítico esa persona no dispone de la creencia de que comer es obsceno de un modo que sea compatible con una conciencia genuina.[11] Por el momento, la creencia sólo está disponible como resultado de una indagación que termina con el testimonio del psicoanalista. Que la persona aún no crea conscientemente que comer es obsceno puede mostrarse, entre otras cosas, en el hecho de que, a pesar de estar dispuesta a autoadscribirse esa creencia mediante una aserción sincera de "Yo creo que comer es obsceno", aún no está dispuesta a expresar espontáneamente esa creencia mediante una aserción sincera de "Comer es obsceno" (Moran 2001, pp. 84–94; Corbí 2010, pp. 333–335). Entonces, 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 (Finkelstein 2003, pp. 115–117).[12]

[11] En otras palabras, aún falta que la persona psicoanalizada transforme ese conocimiento, especialmente el del deseo reprimido que está detrás de esa creencia inconsciente, en una forma genuina de autoconocimiento, a través de la superación de las resistencias a la cura. Como Freud señala: "nuestro saber se habrá convertido entonces también en *su* saber" (1940/1955b, p. 99).

[12] Desde luego, según la distinción de Finkelstein, una creencia puede ser inconsciente de dos maneras: o bien alguien no tiene conocimiento, o conciencia, de su creencia o bien alguien no cree conscientemente algo. Es claro que el sentido

Sin embargo, es posible formular otro tipo de relación necesaria entre tener conocimiento de una creencia y creer conscientemente:

> (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).

Este principio puede considerarse verdadero porque, de entrada, casos como el ejemplo antes examinado no hacen verdadero el antecedente del condicional y, por ende, no constituyen un contraejemplo (a pesar de que el consecuente es falso). Pero justificar el condicional (SC*) es propiamente justificar la idea de que si el antecedente del condicional es verdadero para cierto tipo de creencia, es decir, si existe una relación necesaria, no contingente, entre tener una creencia de tal tipo y saber que uno mismo tiene esa creencia, entonces no queda espacio para la posibilidad de que una creencia del tipo en cuestión sea inconsciente, en el sentido de no creer conscientemente algo.[13] Ciertamente, en el ejemplo de la persona psicoanalizada he estado considerando la distinción de Finkelstein entre que alguien tenga ahora conocimiento, o conciencia, de su creencia y que alguien crea conscientemente algo, en cuyo caso es posible que uno adquiera el conocimiento de una creencia suya que, sin embargo, es o sigue siendo inconsciente. Aunque la creencia ya no es inconsciente en el sentido de no tener ahora conocimiento, o conciencia, de la creencia, sigue siendo inconsciente en el sentido psicológicamente relevante. Pero parece que este sentido tiene aplicación cuando, a pesar de que uno llegue a tener conocimiento, o conciencia, de su creencia, la creencia fue adquirida sin conocimiento de ello y, por tanto, tiene una realidad que se ha constituido psicológicamente al margen de cualquier conocimiento de esa creencia. De manera que en este punto puede definirse una creencia inconsciente, en el sentido de no creer conscientemente algo, como aquella creencia que tiene una realidad constituida psicológicamente (o funcionalmente) al margen de cualquier conocimiento, o conciencia, de esa creencia. En otras palabras: es posible dar una descripción completa del rol funcional de la creencia en la psicología de una persona sin mencionar el conocimiento de

psicológicamente relevante de "creencia inconsciente" es como la negación de creer conscientemente algo.

[13] De entrada, la idea es que el principio representado en (SC*) define, al menos en parte, la conexión entre tener conocimiento de una creencia y creer conscientemente: no es simplemente una generalización psicológica sobre esa conexión.

esa persona al efecto de que ella misma tiene esa creencia.[14] En ese caso, la creencia puede tener una dinámica propia e independiente del conocimiento que uno pudiera adquirir ulteriormente, como el conocimiento adquirido mediante el testimonio del psicoanalista, que es así el conocimiento, o conciencia, de una creencia inconsciente.[15]

Ahora bien, si 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. Sustituyamos ahora "α" por "ella misma Φ" en el condicional (SC*). Como he argumentado que (4) es una consecuencia de (2), resulta que para las creencias de primera persona el antecedente de (SC*) es verdadero:

(A) Si X cree que ella misma Φ, entonces X sabe que ella misma cree que ella misma Φ.

La posesión de una creencia de primera persona no puede ocurrir, a causa de ser instanciado el análogo mental de "yo", sin la posesión del conocimiento de que uno mismo tiene esa creencia. Es decir, existe una relación necesaria, no contingente, entre tener una creencia de primera persona y saber que uno mismo tiene esa creencia. Pero si mi justificación del condicional (SC*), en términos de la definición de creencia inconsciente, en el sentido de no creer conscientemente algo, es correcta, podemos descargar ahora el consecuente de (SC*):

[14] Shoemaker ha defendido en términos funcionalistas la tesis de que existe una relación necesaria, no contingente, o, como él prefiere decirlo, una relación constitutiva (*self-intimation*), entre cualquier creencia que esté disponible (o accesible), no sólo las creencias de primera persona, y una creencia de segundo orden al efecto de que uno mismo tiene esa creencia: no es posible dar una descripción completa del rol funcional de una creencia así disponible en la psicología de una persona sin mencionar la creencia de segundo orden de esa persona al efecto de que ella misma tiene esa creencia (2009).

[15] Por esta razón, la terapia psicoanalítica no consigue la cura, o el restablecimiento de la salud psíquica, sólo como resultado de un conocimiento adquirido mediante el testimonio del psicoanalista. De hecho, Freud señala, con respecto al deseo reprimido que está detrás de una creencia inconsciente, que la adquisición de ese conocimiento no determina el desenlace: "si llevará a que el yo acepte, previo nuevo examen, una exigencia instintiva que hasta el momento había rechazado, o si vuelve a condenarla, esta vez definitivamente" (1940/1955b, p. 100).

(C) Si X cree que ella misma Φ, entonces X cree conscientemente que ella misma Φ.

La conclusión es que, además, existe una relación necesaria, no contingente, entre tener una creencia de primera persona y creer conscientemente.[16]

Pues, si X cree que ella misma Φ, entonces, como la posesión de esa creencia no puede ocurrir sin la posesión del conocimiento de la creencia y, por tanto, la creencia no tiene una realidad constituida psicológicamente al margen de ese conocimiento, la conclusión es que necesariamente la creencia de que ella misma Φ es consciente.

Es importante entender que esta conclusión se aplica a las creencias de primera persona no sólo cuando son estados mentales actuales u ocurrentes, sino también cuando son estados *disposicionales*. Supongamos que alguien, que no está pensando actualmente acerca de su condición de enferma celíaca, tiene la creencia de que ella misma es una enferma celíaca en el sentido de que si alguien le preguntara acerca de su estado de salud o tuviera que tomar una decisión en el almuerzo, actuaría en consecuencia y, además, es muy probable que la creencia pasase a ser un estado mental actual. Pero si, como he argüido, la posesión de una creencia de primera persona no puede ocurrir sin la posesión del conocimiento de que uno mismo tiene esa creencia, entonces, quien cree disposicionalmente que ella misma es una enferma celíaca también conoce disposicionalmente que ella misma tiene esa creencia. Podríamos decir igualmente que, en este caso, la posesión disposicional de una creencia de primera persona no puede ocurrir sin la posesión disposicional del conocimiento de que uno mismo tiene esa creencia. Pues, a pesar de que la creencia sea un estado disposicional, no actual, no puede haber sido adquirida sin conocimiento de ello y, por tanto, no tiene una realidad constituida psicológicamente (o funcionalmente) al margen de ese conocimiento. No es posible describir el rol funcional de esa creencia como un estado disposicional sin mencionar el conocimiento de esa creencia como un estado disposicional. Entonces, de mi definición de

[16] Por otra parte, es obvio que si X cree conscientemente que α, entonces X cree que α. Así que al sustituir "α" por "ella misma Φ", el resultado es éste: si X cree conscientemente que ella misma Φ, entonces X cree que ella misma Φ. De manera que la relación entre tener una creencia de primera persona y creer conscientemente es más estrictamente ésta:

(C*) X cree conscientemente que ella misma Φ si, y sólo si, X cree que ella misma Φ.

creencia inconsciente, en el sentido de no creer conscientemente algo, se sigue ahora que, al creer disposicionalmente que uno mismo es un enfermo celíaco y así conocer disposicionalmente que uno mismo tiene esa creencia, uno cree conscientemente que uno mismo es un enfermo celíaco.

Ciertamente, al creer disposicionalmente que uno mismo es un enfermo celíaco, no puede haber conocimiento, o conciencia, actual de esa creencia. Pero recordemos en este punto que no he dado una definición de creencia inconsciente, en el sentido psicológicamente relevante, en términos de la falta de conocimiento, o conciencia, actual de esa creencia. Como vimos con el ejemplo de la persona psicoanalizada, es posible que una creencia sea inconsciente a pesar de tener conocimiento, o conciencia, actual de esa creencia. Pues, he propuesto que la distinción entre una creencia inconsciente y una creencia consciente consiste en que la creencia tenga o no tenga una realidad *constituida* psicológicamente al margen del conocimiento de esa creencia, respectivamente. Que la creencia sea consciente es cuestión de que la creencia no fuese adquirida al margen de ese conocimiento y, por tanto, esté integrada en la psicología de una persona a través del conocimiento que ella tiene de la creencia, independientemente de que ese conocimiento sea o no actual. Como he argumentado que esta condición es satisfecha por las creencias de primera persona, resulta que las creencias de primera persona como estados disposicionales también son creencias necesariamente conscientes. Al otro extremo del espectro con relación a la persona psicoanalizada, vemos ahora que es posible que una creencia sea consciente a pesar de no tener conocimiento, o conciencia, actual de esa creencia.[17]

Desde luego, cuando una creencia de primera persona es un estado mental actual u ocurrente, tener la creencia entraña tener un conocimiento, o conciencia, actual de esa creencia. En efecto, o bien la creencia es la actualización de una creencia previamente adquirida o bien la creencia es adquirida actualmente. En el primer caso la creencia es la actualización de una creencia cuya posesión disposicional previa no puede ocurrir, a causa de ser instanciado el análogo mental de "yo", sin la posesión disposicional del conocimiento de esa creencia. Pero, entonces, cuando la creencia es actualizada también

[17] Esta propuesta no es tan marginal como pudiera parecer. En este aspecto, Shoemaker sostiene que una creencia es consciente, en un sentido psicológicamente (o funcionalmente) relevante, cuando está *disponible* para guiar el razonamiento y la acción, independientemente de que esa creencia sea o no ocurrente, es decir, haya o no haya sido actualizada (2009, p. 40).

es necesariamente actualizado el conocimiento previo de esa creencia. En el segundo caso es adquirida actualmente una creencia cuya posesión actual no puede ocurrir, a causa de ser instanciado el análogo mental de "yo", sin la posesión actual del conocimiento de esa creencia. Dado que en el caso de la creencia de primera persona que es adquirida actualmente, que no es la actualización de una creencia disposicional previa, la creencia es necesariamente consciente, en el sentido de creer conscientemente algo, porque no fue adquirida al margen del conocimiento, o conciencia, *actual* de esa creencia, se sigue que en este caso, mientras la creencia de primera persona es actual, creer conscientemente *es* tener un conocimiento, o conciencia, actual de la creencia.[18]

II

He argumentado hasta aquí que ninguna creencia de primera persona puede ser inconsciente en el sentido psicológicamente relevante. Esta conclusión amenaza la costumbre psicoanalítica de describir una creencia inconsciente como si tuviera un contenido de primera persona, a pesar de que la metapsicología analítica no es compatible con ello. Consideremos, como un ejemplo paradigmático, el caso, presentado por el propio Freud, de aquella mujer de clase alta que tendría la creencia inconsciente de que su marido *le* es infiel, producida como resultado de un mecanismo de desplazamiento que, a partir de un deseo reprimido hacia su hijo adoptivo, genera unos celos ilusorios (Corbí 2010). Si mi argumentación anterior es correcta, aquella mujer no podía tener la creencia inconsciente de que ella misma estaba siendo traicionada por su marido. Pero esta consecuencia se sigue también de la metapsicología analítica, la tópica del psiquismo, según la cual los estados mentales inconscientes no son estados del *yo* sino estados del *ello* (Freud 1923/1955a), de manera que cuando tienen un contenido reflexivo, esto es, cuando tienen un contenido acerca de su poseedor, no puede ser un contenido de primera persona; excepto que el ello sea ilegítimamente personificado, como de hecho es una mala costumbre psicoanalítica.[19]

[18] Precisamente cuando la creencia deja de ser actual pero se conserva como un estado disposicional, creer conscientemente ya no es tener un conocimiento, o conciencia, actual de la creencia.

[19] Como señalé en la nota 2, podría argumentarse más ambiciosamente que ningún estado mental inconsciente, como los deseos reprimidos, puede tener un contenido de primera persona. Puede mostrarse que es así con respecto a los deseos de primera persona mediante una formulación de (2), (3), (4), (SC*) (A) y (C) equivalente a la

La conclusión general de que ninguna creencia de primera persona puede ser inconsciente es especialmente relevante con respecto a las creencias de primera persona con un contenido psicológico, como las creencias acerca de nuestros estados mentales actuales cuando son *conscientes*. Esta consecuencia constituye una importante objeción contra un cierto tipo de teoría de orden superior de la conciencia según la cual la naturaleza consciente de un estado mental actual como, por ejemplo, una sensación de dolor consiste precisamente en tener un pensamiento o una creencia[20] de orden superior con el contenido de primera persona: que uno mismo siente dolor (Rosenthal 1997 y 2005). Tengamos en cuenta, además, que el pensamiento de orden superior también es un estado mental actual u ocurrente. La teoría establece así, en otras palabras, que un estado mental actual es consciente cuando mediante un pensamiento de orden superior acerca de ese estado uno tiene conciencia actual de ese estado o, más precisamente en términos de primera persona, uno tiene conciencia actual de uno mismo como encontrándose en ese estado. Más aún, téngase en cuenta que el pensamiento de orden superior es actual sin ser la actualización de una creencia disposicional previa, pues, al ser un pensamiento acerca de un estado mental *actual*, no puede ser la actualización de una creencia disposicional previa acerca de ese estado. Pero mostré que cuando una creencia de primera persona es actual u ocurrente sin ser la actualización de una creencia disposicional previa, creer conscientemente *es* tener un conocimiento, o conciencia, actual de la creencia. Puesto que, como cualquier otra creencia o pensamiento de primera persona, un pensamiento de orden superior es necesariamente consciente en el sentido de pensar conscientemente algo, como he estado arguyendo, se sigue ahora que también es necesariamente consciente en el sentido de tener conciencia actual de ese pensamiento.

Ahora bien, una de las ventajas de la teoría de orden superior era explicar la naturaleza consciente de un estado mental en términos de

formulación que he presentado para las creencias de primera persona. Sin embargo, insisto en que no está en juego un cuestionamiento del concepto psicoanalítico de inconsciente sino, más bien, de la personificación del inconsciente que la propia metapsicología analítica niega.

[20] Rosenthal prefiere hablar de pensamientos más que de creencias porque se trata de estados actuales u ocurrentes, pero no caracteriza un pensamiento como un estado intencional con un contenido proposicional, algo que es común tanto a las creencias como a los deseos o intenciones, sino como un estado intencional que es el análogo mental de una aserción, como las creencias solamente (1997, p. 742). A continuación daré por supuesto este sentido.

la posesión de un pensamiento de orden superior que fuera inconsciente, en el sentido de no tener conciencia actual de ese pensamiento. De otro modo, la naturaleza consciente de ese pensamiento tendría que explicarse, por aplicación de la propia teoría, en términos de un pensamiento de un orden más alto aún, y entonces podría generarse un regreso al infinito. Por ejemplo, se trataría de un pensamiento de primera persona con el contenido: que uno mismo piensa que uno mismo siente dolor. Pero si cualquier pensamiento de primera persona que es actual sin ser la actualización de una creencia disposicional previa es necesariamente consciente, en el sentido de tener conciencia actual de ese pensamiento, ese pensamiento de un orden más alto aún también sería consciente en ese sentido, en cuyo caso la naturaleza consciente de ese pensamiento tendría que explicarse, por aplicación de la propia teoría, en términos de un pensamiento de un orden más alto aún, y así sucesivamente. Esto significa que la teoría de Rosenthal debe abandonarse: ¡no puede ser que un estado mental consciente, como una sensación de dolor, tenga que ir acompañado actualmente por un número infinito de estados mentales conscientes![21]

Como veremos a continuación, Rosenthal advierte otro peligro que para su teoría de orden superior supone cierta concepción del análogo mental de "yo" en un pensamiento de primera persona, que entonces sería necesariamente consciente. Por eso, trata de distinguir entre una autorreferencia esencialmente indéxica según la cual el análogo mental de "yo" en un pensamiento de primera persona refiere a uno mismo, el pensador, y una autorreferencia más robusta según la cual el análogo mental de "yo" en un pensamiento de primera persona refiere a uno mismo, el pensador, *como el pensador de ese pensamiento*, o, por decirlo en los términos que introduje desde el principio, refiere a uno mismo como uno mismo.[22] Pero, desde luego,

[21] Esta objeción es válida porque es esencial al planteamiento de Rosenthal que un estado mental y el pensamiento de orden superior acerca de ese estado, así como cualquier pensamiento de un orden más alto aún, sean estados mentales numéricamente distintos. De hecho, la teoría de orden superior establece que se trata de estados mentales que están en una relación contingente (o extrínseca) y que entonces, al estar en una relación contingente con un pensamiento de un orden más alto aún, un pensamiento de orden superior no es necesariamente consciente (1997, pp. 735–743; 2005, pp. 26–31).

[22] En el análisis que realicé en la primera parte del artículo no me comprometí con la tesis de que la referencia de primera persona en un pensamiento consiste en referirse, mediante una descripción, a uno mismo como el pensador de ese pensamiento, es decir, como quien hace la referencia de primera persona. Fui neutral con respecto al modo de presentación de uno mismo *como uno mismo*, pues defendí que deberíamos aceptar (3) y (4) como un dato que tiene que ser explicado por cualquier

una autorreferencia esencialmente indéxica debe distinguirse de cualquier otra forma de referirse accidentalmente a uno mismo, como cuando Edipo se refiere a él mismo sin saberlo pensando en el hijo de Layo. La idea es que, además de referirse a uno mismo, una autorreferencia esencialmente indéxica en un pensamiento de primera persona requiere estar *dispuesto* a describirse uno mismo como el pensador de ese pensamiento. No requiere describirse de hecho, en ese pensamiento de primera persona, como el pensador de ese pensamiento sino estar dispuesto a describirse, en otro pensamiento acerca de ese pensamiento de primera persona, como el pensador de ese pensamiento: "Pero siempre que tengo un pensamiento de primera persona de que yo soy *F* [Yo Φ], tener ese pensamiento me dispone a tener otro pensamiento que identifica [describe] el individuo acerca del cual es ese pensamiento como el pensador de ese pensamiento" (Rosenthal 2004, p. 167; 2012, pp. 30–31).

De manera que, respecto a cualquier pensamiento de primera persona, referirse de un modo esencialmente indéxico a uno mismo es estar dispuesto a tener otro pensamiento acerca de ese pensamiento de primera persona. Ahora bien, un pensamiento de orden superior acerca de un estado mental es un pensamiento de primera persona y, por tanto, el análogo mental de "yo" en ese pensamiento hace una autorreferencia esencialmente indéxica. Entonces, respecto a un pensamiento de orden superior acerca de un estado mental, referirse de un modo esencialmente indéxico a uno mismo es estar dispuesto a tener un pensamiento de un orden más alto aún acerca de ese pensamiento.

Pero, ¿en qué sentido la distinción entre una autorreferencia esencialmente indéxica y una autorreferencia más robusta trata de salvar la teoría de orden superior de la conciencia? Rosenthal sostiene que si el análogo mental de "yo" en un pensamiento de primera persona refiriera a uno mismo como uno mismo, como el pensador de ese pensamiento, entonces el pensamiento sería necesariamente consciente. De la misma forma, una autorreferencia más robusta entrañaría

teoría sobre el significado y la referencia del pronombre personal "yo" que reconozca la necesidad de un modo de presentación con esas características, independientemente de cómo se conciba ese modo. Me parece que, en efecto, deberíamos aceptar (3) y (4) tanto si ese modo de presentación se concibe en los términos reflexivistas de referirse mediante una descripción a uno mismo como el pensador del pensamiento, como si, por ejemplo, se concibe fregeanamente como un modo de presentación "especial y primitivo" (Recanati 2007, pp. 169–170). En este último caso, uno sabe de una manera especial y primitiva que uno mismo es el pensador del pensamiento.

que un pensamiento de orden superior acerca de un estado mental fuera necesariamente consciente:

> Es importante para el modelo que cuando un pensamiento refiere a uno mismo de este modo esencialmente indéxico, su contenido no describa al individuo a que refiere como el pensador del pensamiento. Si un pensamiento de primera persona esencialmente indéxico describiera al individuo acerca del cual es como el pensador de ese pensamiento, simplemente tener ese pensamiento haría que uno fuese consciente de tenerlo. Y, como los pensamientos de orden superior son pensamientos de primera persona esencialmente indéxicos, uno no podría tener un pensamiento de orden superior sin ser consciente de uno mismo como teniéndolo. Pero somos totalmente inconscientes de la mayor parte de nuestros pensamientos de orden superior. (Rosenthal 2004, p. 167; 2012, p. 30)

El problema no es que haya pensamientos de orden superior que sean conscientes. Pero una de las ventajas de su teoría era distinguir entre estados mentales conscientes y estados mentales que, además, están bajo introspección, dado que parece que la mayor parte de nuestros estados mentales conscientes no son introspeccionados. De acuerdo con Rosenthal, esto requiere distinguir a su vez entre pensamientos de orden superior conscientes e inconscientes, y así explicar la conciencia (no introspectiva) de un estado mental en términos de un pensamiento de orden superior que es inconsciente. En efecto, la introspección, que, a diferencia de la mera conciencia de un estado mental, consiste en ser consciente de ese estado de un modo reflexivo y atento, equivale, según Rosenthal, a tener un pensamiento de orden superior que es consciente.[23] Por tanto, en términos de la propia teoría de orden superior, la introspección de un estado mental consiste en tener un pensamiento de un orden más alto aún mediante el cual uno es consciente de uno mismo como teniendo el pensamiento de orden superior mediante el cual uno es consciente de uno mismo como encontrándose en ese estado (Rosenthal 1997, pp. 745–746; 2005, pp. 107–113). Pero si un pensamiento de orden superior describiera al individuo acerca del cual es ese pensamiento como el pensador de ese pensamiento, sería necesariamente consciente: mediante un pensamiento de orden superior uno sería consciente de uno mismo como teniendo *ese* pensamiento de orden superior. Entonces, todos

[23] En este sentido, el hecho de que la mayor parte de los pensamientos de orden superior son inconscientes *es* el hecho de que la mayor parte de los estados mentales conscientes no están bajo introspección.

los estados mentales conscientes serían necesariamente estados bajo introspección. Pues, ser consciente de un estado mental de un modo reflexivo y atento consistiría simplemente en tener el pensamiento de orden superior mediante el cual uno es consciente de uno mismo como encontrándose en ese estado.

Así que Rosenthal está de acuerdo con la tesis de que si el análogo mental de "yo" en un pensamiento de primera persona refiere a uno mismo como uno mismo, entonces ese pensamiento es necesariamente consciente. Pero, desde luego, abandona este tipo de autorreferencia más robusta para salvar la distinción, que su teoría establece, entre estados mentales conscientes y estados mentales que, además, están bajo introspección. Sin embargo, hay al menos una razón importante por la que la idea de una autorreferencia esencialmente indéxica, distinta de esa autorreferencia más robusta, no es satisfactoria. Rosenthal sostiene que una autorreferencia esencialmente indéxica a uno mismo en el contenido de un pensamiento de primera persona de la forma "Yo Φ" consiste en estar dispuesto a tener un pensamiento de orden superior acerca de ese pensamiento. Ahora bien, Rosenthal da por hecho que la disposición será actualizada cuando el individuo al que refiere *ese pensamiento de primera persona* es descrito como el pensador de ese pensamiento mediante un pensamiento de orden superior de la forma "Yo pienso que yo Φ". El problema es que, según su propio planteamiento, no hay garantías de que sea así. Un pensamiento de orden superior, como cualquier pensamiento de primera persona, refiere a uno mismo, el pensador de ese pensamiento de orden superior. En ese caso, un pensamiento de orden superior de la forma "Yo pienso que yo Φ" describe al individuo al que refiere *ese pensamiento de orden superior*, mediante la primera instanciación de "yo", como el pensador de un pensamiento de primera persona de la forma "Yo Φ". Entonces, una autorreferencia esencialmente indéxica a uno mismo en el contenido de un pensamiento de primera persona consiste realmente en estar dispuesto a tener un pensamiento de orden superior que describa al pensador de ese pensamiento de orden superior como el pensador de ese pensamiento de primera persona. Ciertamente, en condiciones normales, tanto el pensamiento de primera persona como el pensamiento de orden superior acerca de ese pensamiento refieren al mismo individuo y, por tanto, referirse al pensador del pensamiento de orden superior como el pensador del pensamiento de primera persona *es* referirse al pensador de ese pensamiento de primera persona como tal.

Sin embargo, Rosenthal afirma que un pensamiento de orden superior acerca de un estado mental no es inmune a un error de iden-

tificación: "Aunque sea improbable que uno esté en lo correcto al pensar que alguien se encuentra en un estado particular pero se equivoque en que el individuo en ese estado es uno mismo, tal error no es imposible" (2004, p. 170; 2005, pp. 355–356). Entonces, la primera instanciación del análogo mental de "yo" en un pensamiento de orden superior de la forma "Yo pienso que yo Φ" no es inmune a un error de identificación. Es posible que un pensamiento de orden superior acerca de un pensamiento de primera persona, que lo describe a uno mismo, el pensador de ese pensamiento de orden superior, como el pensador de ese pensamiento de primera persona, sea falso a ese respecto. Esto significa que tener un pensamiento de orden superior que describe al pensador de ese pensamiento de orden superior como el pensador del pensamiento de primera persona no garantiza que efectivamente sea el pensador de ese pensamiento de primera persona quien es descrito como tal. Es importante señalar que Rosenthal niega que haya inmunidad a un error de identificación no sólo cuando el pensamiento de orden superior acerca de un estado mental está basado en una inferencia consciente, por ejemplo, a partir de la observación de la conducta, sino también cuando se trata de un pensamiento no inferencial, en el sentido de que el pensamiento de orden superior parece directo o inmediato con respecto al estado mental, como es propio de un pensamiento de orden superior que es la actualización de una disposición previa a tener ese pensamiento.[24] Parece así que una disposición a tener un pensamiento de orden superior que describa al pensador de ese pensamiento de orden superior como el pensador del pensamiento de primera persona puede ser actualizada sin que efectivamente sea el pensador de ese pensamiento de primera persona quien es descrito como tal. Pero no tiene sentido que una autorreferencia esencialmente indéxica a uno mismo en el contenido de un pensamiento de primera persona esté basada en una disposición cuya actualización puede errar de esa manera.

Concluyo que es mejor suponer que el análogo mental de "yo" en un pensamiento de primera persona refiere a uno mismo como uno mismo, en cuyo caso, como Rosenthal reconoce, los pensamientos de primera persona son necesariamente conscientes. Por lo tanto, debe proponerse, so pena de una amenaza de regreso al infinito, una teoría

[24] Como ejemplo de ello, considera los casos donde hay un desorden psicológico de identidad disociativa (Rosenthal 2012, pp. 42–44). Precisamente son casos en los que puede ocurrir que un *alter ego* actualice la disposición a tener un pensamiento de orden superior, acerca de un pensamiento de primera persona, mediante el cual se describe erróneamente a sí mismo como el pensador de ese pensamiento de primera persona.

alternativa a la teoría de orden superior de la conciencia según la cual la naturaleza consciente de un estado mental como un dolor no requiera tener un pensamiento de primera persona acerca de ese estado.

III

Voy a argumentar ahora que la tesis de que no puede haber creencias de primera persona que sean inconscientes permite dar una explicación precisa del sinsentido o irracionalidad propio de la paradoja de Eroom. Esta paradoja ha sido introducida en la literatura por David Finkelstein invirtiendo el nombre de la paradoja de Moore, que, como es sabido, consiste en el sinsentido de realizar una aserción o tener una creencia con un contenido de la forma "*p* pero yo no creo que *p*" o, lo que es más relevante en este contexto, de la forma "no *p* pero yo creo que *p*". La paradoja de Eroom consiste en el sinsentido de realizar una aserción o tener una creencia con un contenido de la forma (Finkelstein 2003, pp. 118–119; Gallois 2007):

(5) *p* y yo creo inconscientemente que *p*.

Como ocurre con la paradoja de Moore, resulta que ambos miembros de la conjunción pueden ser verdaderos a la vez sin que haya una contradicción, y de hecho lo son siempre que alguien tiene una creencia inconsciente que es verdadera, pero parece intuitivamente que hay algo absurdo o irracional en realizar una aserción o tener una creencia con ese contenido conjuntivo.[25] Ahora bien, un contenido conjuntivo de esta forma es parcialmente, debido al segundo miembro de la conjunción, un contenido de primera persona y, en ese caso, a nivel psicológico de la creencia la paradoja de Eroom

[25] Finkelstein da una explicación de la irracionalidad de realizar una aserción, pero no de tener una creencia, con el contenido de la paradoja de Eroom. Según el expresivismo que defiende, una creencia consciente de que *p* es aquella que puede ser expresada o manifestada mediante una autoadscripción de la forma "Yo creo que *p*". De manera que si alguien aseverara algo de la forma "*p* y yo creo inconscientemente que *p*", su aserción sería verdadera si, y sólo si, *p* y él no tuviera la capacidad de expresar la creencia de que *p* mediante una autoadscripción de esa forma. Pero es absurdo que alguien pueda aseverar sinceramente que *p* y, sin embargo, no pueda expresar la creencia de que *p* mediante una autoadscripción. Pues, cualquiera que es capaz de expresar su creencia de que *p* mediante una aserción de "*p*" también es capaz de expresarla mediante una aserción de "Yo creo que *p*" (2003, p. 121).

consiste en el sinsentido de tener cierta creencia de primera persona.[26] Pero, como voy a argumentar, el hecho de que una creencia de primera persona sea necesariamente consciente explica precisamente por qué es un sinsentido tener una creencia de primera persona con un contenido de la forma descrita en (5).[27]

De entrada, la adscripción de una creencia de primera persona con el contenido de la paradoja de Eroom, una creencia de Eroom, es de la forma:

(6) *X* cree que (*p* y ella misma cree inconscientemente que *p*).

Como señalé en la primera sección, el cuasi-indéxico "ella misma" es la traducción a contextos indirectos como (6) del pronombre personal "yo", usado en cualquier sustitución de (5) para expresar directamente el contenido de una creencia de Eroom. Entonces, la verdad de una adscripción como (6) se caracteriza porque, como en esa creencia *X* se refiere con éxito a ella misma *como ella misma*, es imposible que la persona referida por "*X*" en la adscripción (fuera de la cláusula-que), quien es el sujeto de la creencia, 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. De manera que de (6) se sigue:

(7) *X* sabe que la creencia de que (*p* y alguien cree inconscientemente que *p*) es acerca de quien cree que (*p* y alguien cree inconscientemente que *p*), es decir, ella misma.

Pero, como según esto *X* sabe que ella misma es tanto el objeto como el sujeto de la creencia de que (*p* y alguien cree inconscientemente

[26] La tesis de que una creencia de Eroom es una creencia de primera persona, a pesar de no tener un contenido neto de la forma "Yo Φ", se justifica en el hecho de que, a causa de ser instanciado el análogo mental de "yo", una creencia de Eroom tiene las propiedades de cualquier creencia de primera persona de la forma "Yo Φ", como sostengo a continuación. Entonces, una creencia de primera persona debería redefinirse como aquella creencia en la que es instanciado el análogo mental de "yo". A partir de esta consideración habría que reformular los puntos (1), (2), (3), (4), (A) y (C) de la primera parte. Pero me parece que, a efectos de mantener un tratamiento estándar de las creencias de primera persona, es mejor dejar la primera parte como está.

[27] También se explicaría así por qué es un sinsentido realizar una aserción con un contenido de la forma descrita en (5), según el principio de Shoemaker: lo que puede creerse coherentemente (o racionalmente) constriñe lo que puede aseverarse coherentemente. Pues, una aserción es una manifestación de la creencia y, por tanto, una aserción que trata de manifestar una creencia incoherente también será incoherente (Green y Williams 2007, p. 12).

que p), resulta que "ella misma" puede ocupar la posición tanto del sujeto de la creencia (fuera de la cláusula-que subordinada) como del objeto de la creencia (dentro de la cláusula-que subordinada) en la siguiente adscripción de ese conocimiento:

(8) X sabe que ella misma cree que (p y ella misma cree inconscientemente que p).

Ahora bien, tengamos en cuenta el condicional (SC*), la relación necesaria entre tener conocimiento de una creencia y creer conscientemente, que establecimos anteriormente, según la cual no es posible que el antecedente del condicional sea verdadero para un cierto tipo de creencia y que, sin embargo, una creencia de tal tipo sea inconsciente en el sentido psicológicamente relevante. Pues, si la posesión de una creencia de tal tipo no puede ocurrir sin la posesión del conocimiento de que uno mismo tiene la creencia y, por tanto, la creencia no tiene una realidad psicológicamente constituida al margen de ese conocimiento, la conclusión es que necesariamente la creencia es consciente. Sustituyamos ahora "α" por "p y ella misma cree inconscientemente que p" en el condicional (SC*). En la primera sección argumenté también que el antecedente de (SC*) es verdadero para las creencias de primera persona. La posesión de una creencia de primera persona no puede ocurrir, a causa de ser instanciado el análogo mental de "yo", sin la posesión del conocimiento de que uno mismo tiene esa creencia. Esto es manifiesto precisamente en el hecho de que, como hemos visto, (8) es una consecuencia de (6):

(9) Si X cree que (p y ella misma cree inconscientemente que p), entonces X sabe que ella misma cree que (p y ella misma cree inconscientemente que p).

De manera que si mi justificación del condicional (SC*), en términos de la definición de creencia inconsciente, en el sentido de no creer conscientemente algo, fue correcta, podemos descargar ahora el consecuente de (SC*):

(10) Si X cree que (p y ella misma cree inconscientemente que p), entonces X cree conscientemente que (p y ella misma cree inconscientemente que p).

Pues, si X cree que (p y ella misma cree inconscientemente que p), entonces, como la posesión de esa creencia no puede ocurrir sin la posesión del conocimiento de la creencia y, por tanto, la creencia

no tiene una realidad constituida psicológicamente al margen de ese conocimiento, la conclusión es que necesariamente la creencia de que (p y ella misma cree inconscientemente que p) es consciente.

Desde luego, de (6) y de (10) obtenemos finalmente:

(11) X cree conscientemente que (p y ella misma cree inconscientemente que p).

Pero parece que la creencia consciente se distribuye de manera trivial sobre la conjunción: si X cree conscientemente que (p y q), entonces X cree conscientemente que p y X cree conscientemente que q (Williams 2006, p. 394).[28] De modo que de (11) obtenemos: X cree conscientemente que p y X cree conscientemente que ella misma cree inconscientemente que p.[29] Alguien podría señalar ahora que si X tuviera la mente consciente dividida en centros de conciencia[30] incomunicados entre sí, entonces podría racionalmente creer conscientemente que p y creer conscientemente que ella misma cree

[28] Williams sostiene que, a diferencia de lo que ocurre con la creencia *simpliciter*, también es algo trivial que la creencia consciente se recoge bajo la conjunción: si X cree conscientemente que p y X cree conscientemente que q, entonces X cree conscientemente que (p y q). Sin embargo, me parece que esta tesis requeriría una línea independiente de argumentación. Pues, como señalo a continuación, podemos pensar que una creencia consciente de que p y una creencia consciente de que q pertenecen a distintos centros de conciencia incomunicados entre sí, en cuyo caso no se sigue que haya una creencia consciente de que (p y q).

[29] Téngase en cuenta que puede reconstruirse el caso de la persona psicoanalizada que llega a saber que tiene la creencia inconsciente de que comer es obsceno como el caso de alguien que, en una primera fase, llega a creer conscientemente que ella misma cree inconscientemente que comer es obsceno. Con todo, aún no cree conscientemente que comer es obsceno. No parece que sea irracional creer conscientemente que uno mismo cree inconscientemente que p y, sin embargo, no creer conscientemente que p.

[30] En esta primera explicación estoy suponiendo que la creencia de Eroom es actual u ocurrente sin ser la actualización de una creencia disposicional previa. Como señalé en la primera parte del artículo, sólo en ese caso resulta que creer conscientemente es tener un conocimiento, o conciencia, actual de esa creencia. Además, parece que sólo cuando hay conciencia actual tiene sentido hablar de centros de conciencia o de creencias co-conscientes, como hago a continuación. Pero, aunque en las discusiones tanto de la paradoja de Moore como de la de Eroom se da por supuesto que en efecto se trata de creencias actuales, puede plantearse que esta explicación no es correcta si la creencia de Eroom es concebida como un estado disposicional. Sin embargo, podría darse una interpretación disposicional de nociones como "centro de conciencia" o "co-consciente". Por ejemplo, podría decirse que un centro de conciencia se individúa en términos de un conjunto de creencias, y otros estados, cuya posesión disposicional no puede ocurrir sin la posesión disposicional del conocimiento de esas creencias. Ahora bien, téngase en cuenta que, independientemente de que esta estrategia funcione o no, la segunda explicación de la paradoja

inconscientemente que *p*. Pues, es racional creer conscientemente que *p* y a la vez creer conscientemente que la creencia de que *p*, que no pertenece al mismo centro de conciencia que esa otra creencia acerca de ella, es inconsciente. Sin embargo, es evidente que (11) representa la pertenencia de ambas creencias conscientes al mismo centro de conciencia, o por decirlo kantianamente, a la misma *unidad de la conciencia*: si X cree conscientemente que (*p* y ella misma cree inconscientemente que *p*), no puede ocurrir que la creencia consciente de que *p* no pertenezca al mismo centro de conciencia que la creencia consciente de que ella misma cree inconscientemente que *p*. En otras palabras, no puede ocurrir que ambas creencias conscientes no sean co-conscientes.[31] Pero parece que es irracional creer conscientemente que *p* y a la vez creer conscientemente que la creencia de que *p*, que pertenece a la misma unidad de la conciencia que esa otra creencia acerca de ella, es inconsciente. Es irracional tener creencias co-conscientes tales que una de ellas establece que la otra es una creencia inconsciente. Esto significa que, desde el punto de vista de la racionalidad de la creencia, *X* debería suspender la creencia consciente de que ella misma cree inconscientemente que *p*.

Así, la unidad de la conciencia representada en (11) explica, en primer lugar, que no pueda ocurrir que la mente consciente de *X* esté dividida en centros de conciencia incomunicados entre sí, un centro al que pertenece la creencia consciente de que *p* y un centro al que pertenece la creencia consciente de que ella misma cree que *p*, que entonces podría racionalmente ser la creencia consciente de que ella misma cree *inconscientemente* que *p*. Pero resulta que de la unidad de la conciencia se sigue, en segundo lugar, que cualquier creencia consciente acerca de la creencia de que *p* que *X* tenga cuando a la vez tiene la creencia consciente de que *p* ha de ser, desde el punto de vista de la racionalidad de la creencia, la creencia consciente de que ella misma cree *conscientemente* que *p*. De manera que (11) representa una unidad de la conciencia que, a causa del contenido de las creencias conscientes que *X* tendría, no puede darse racionalmente.

que voy a presentar puede aplicarse a la creencia de Eroom concebida tanto actual como disposicionalmente.

[31] A este respecto, podría enunciarse el siguiente principio, que es una verdad trivial:

(UC) Si *X* cree conscientemente que (*p* y *q*), entonces la creencia consciente de que *p* y la creencia consciente de que *q* son co-conscientes.

La conclusión es que es irracional que X crea conscientemente que (p y ella misma cree inconscientemente que p).[32]

Aunque esta explicación en términos de la unidad de la conciencia está basada en el contenido de primera persona de la creencia de Eroom, que implica que esa creencia es necesariamente consciente, puede mostrarse más explícitamente el rol del análogo mental de "yo" en la irracionalidad de esa creencia. De entrada, parece que la creencia, sea o no sea consciente, se distribuye trivialmente sobre la conjunción: si X cree que (p y q), entonces X cree que p y X cree que q. De modo que de (6) obtenemos: X cree que p y X cree que ella misma cree inconscientemente que p (Williams 2006, p. 393). Entonces, de (7) y de la distribución de la creencia sobre cada miembro de la conjunción, de hecho sobre el segundo miembro y luego sobre el primer miembro, se sigue:

(7*) X sabe que la creencia de que alguien cree inconscientemente que p es acerca de quien cree que p, es decir, ella misma.

Ahora bien, del mismo modo que (7), a causa del contenido de primera persona de la creencia de Eroom, es una consecuencia de (6), resulta que de (11) se sigue:

(12) X sabe que la creencia consciente de que (p y alguien cree inconscientemente que p) es acerca de quien cree conscientemente que (p y alguien cree inconscientemente que p), es decir, ella misma.

Pero, como hemos visto, parece que, al igual que la creencia, la creencia consciente se distribuye trivialmente sobre la conjunción. Entonces, de (12) y de la distribución de la creencia consciente sobre la conjunción, de hecho sobre el segundo miembro y luego sobre el primer miembro, se sigue:

(12*) X sabe que la creencia consciente de que alguien cree inconscientemente que p es acerca de quien cree conscientemente que p, es decir, ella misma.

[32] Ni esta explicación basada en la unidad de la conciencia ni la explicación que propongo a continuación consisten en establecer que alguien con una creencia de Eroom tendría una creencia contradictoria, una creencia cuyo contenido es una contradicción, o tendría dos creencias inconsistentes, dos creencias cuyos contenidos son contradictorios entre sí. Primero, la creencia consciente de que (p y uno mismo cree inconscientemente que p) no es, obviamente, una creencia contradictoria. Segundo, la creencia consciente de que p y la creencia consciente de que uno mismo cree inconscientemente que p tampoco son creencias inconsistentes. Pero esto no significa que no haya otros tipos de irracionalidad, como los considerados en el texto.

Así que X sabe que ella misma es tanto el objeto de la creencia consciente de que alguien cree inconscientemente que p como el sujeto de la creencia consciente de que p. En otras palabras, X sabe que la persona de quien cree conscientemente que cree inconscientemente que p es la misma persona, ella misma, que cree conscientemente que p. La relevancia explicativa de la primera persona, del análogo mental de "yo" en la creencia de Eroom, está en lo siguiente: del hecho de que X crea conscientemente que ella misma cree inconscientemente que p, que es (la parte de) la creencia con un contenido de primera persona, se sigue que X sabe que la creencia consciente de que alguien cree inconscientemente que p es sobre quien cree conscientemente que p, es decir, ella misma. Pero claramente es irracional que alguien tenga el conocimiento descrito en (12*). Es irracional, por tanto, que X crea conscientemente que (p y ella misma cree inconscientemente que p).[33,34]

BIBLIOGRAFÍA

Bermúdez, J.L., 1998, *The Paradox of Self-Consciousness*, The MIT Press, Cambridge, Mass.

Castañeda, H.N., 1966/1999, "'He': A Study in the Logic of Self-Consciousness", en H.N. Castañeda, *The Phenomeno-Logic of the I. Essays on Self-Consciousness*, Indiana University Press, Bloomington, pp. 35–60.

———, 1989, "The Reflexitivity of Self-Consciousness: Sameness/Identity, Data for Artificial Intelligence", *Philosophical Topics*, vol. 17, no. 1, pp. 27–58.

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.

[33] Téngase en cuenta que la irracionalidad de la creencia de Eroom no puede obtenerse directamente de (7*). Pues (7*) no excluye que la creencia de que p sea inconsciente. Pero, en ese caso, no habría irracionalidad alguna: X sabe que la creencia de que alguien cree inconscientemente que p es sobre quien cree *inconscientemente* que p, es decir, ella misma. Recordemos que la inferencia de (6)–(7) a (11)–(12) depende de la tesis de que las creencias de primera persona son necesariamente creencias conscientes, y es esa tesis, por tanto, la que establece, a través de (12*), la irracionalidad de la creencia de Eroom.

[34] Agradezco el apoyo financiero otorgado por el gobierno de Chile (CONICYT) mediante el Proyecto FONDECYT Regular 2014 no. 1140395. Por otra parte, estoy en deuda con los árbitros de esta revista, quienes con sus observaciones y sugerencias me llevaron a hacer cambios que, sin duda, han mejorado la calidad del artículo. Por último, tengo una deuda especial con Carola Yong Sakanishi.

Corazza, E., 2004, *Reflecting the Mind: Indexicality and Quasi-Indexicality*, Clarendon Press, Oxford.
Corbí, J.E., 2010, "First-Person Authority and Self-Knowledge as an Achievement", *European Journal of Philosophy*, vol. 18, no. 3, pp. 325–362.
Ezcurdia, M., 2001, "Thinking about Myself", en A. Brook (comp.), *Self-Reference and Self-Awareness*, John Benjamins Publishing, Amsterdam, pp. 179–203.
Finkelstein, D., 2003, *Expression and the Inner*, Harvard University Press, Harvard.
Freud, S., 1923/1955a, "El Yo y el Ello", en *Obras completas*, vol. IX, trad. L. Rosenthal, Santiago Rueda, Buenos Aires, pp. 193–237.
———, 1940/1955b, "Compendio del psicoanálisis", en *Obras completas*, vol. XXI, trad. L. Rosenthal, Santiago Rueda, Buenos Aires, pp. 67–126.
Gallois, A., 2007, "Consciousness, Reasons, and Moore's Paradox", en M. Green y J.N. Williams 2007, pp. 165–188.
García-Carpintero, M., 2013, "Self-Conception: Sosa on *De Se* Thought", en J. Turri (comp.), *Virtuous Thoughts: The Philosophy of Ernest Sosa*, Springer, Dordrecht, pp. 73–99.
Geach, P.T., 1957, "On Beliefs about Oneself", *Analysis*, vol. 18, no. 1, pp. 23–24.
Green, M. y J.N. Williams (comps.), 2007, *Moore's Paradox. New Essays on Belief, Rationality, and the First Person*, Cambridge University Press, Cambridge.
Lewis, D., 1979, "Attitudes *De Dicto* and *De Se*", *Philosophical Review*, vol. 88, pp. 513–543.
Moran, R., 2001, *Authority and Estrangement. An Essay on Self-Knowledge*, Princeton University Press, Princeton.
Morgan, D., 2012, "Immunity to Error through Misidentification: What Does It tell Us About the *De Se*?", en S. Prosser y F. Recanati (comps.), *Immunity to Error through Misidentification*, Cambridge University Press, Cambridge, pp. 104–123.
O'Brien, L., 2007, *Self-Knowing Agents*, Oxford University Press, Oxford.
Recanati, F., 2009, "*De re* and *De se*", *Dialectica*, vol. 63, no. 3, pp. 249–269.
———, 2007, *Perspectival Thought. A Plea for (Moderate) Relativism*, Oxford University Press, Oxford.
Rosenthal, D.M., 2012, "Awareness and Identification of Self", en J. Liu y J. Perry (comps.), *Consciousness and the Self*, Cambridge University Press, Cambridge, pp. 22–50.
———, 2005, *Consciousness and Mind*, Oxford University Press, Oxford.
———, 2004, "Being Conscious of Ourselves", *The Monist*, vol. 87, no. 2, pp. 159–181.
———, 1997, "A Theory of Consciousness", en N. Block, O. Flanagan y G. Güzeldere (comps.), *The Nature of Consciousness: Philosophical Debates*, The MIT Press, Cambridge, Mass., pp. 729–753.

Rovane, C., 1987, "The Epistemology of First-Person Reference", *The Journal of Philosophy*, vol. 84, no. 3, march, pp. 147–167.
Shoemaker, S., 2009, "Self-Intimation and Second Order Belief", *Erkenntnis*, vol. 71, no. 1, pp. 35–51.
———, 1968/1994, "Self-Reference and Self-Awareness", en Q. Cassam (comp.), *Self-Knowledge*, Oxford University Press, Oxford, pp. 80–93.
Williams, J.N., 2006, "Moore's Paradox and Conscious Belief", *Philosophical Studies*, vol. 127, no. 3, pp. 383–414.

Recibido el 27 de agosto de 2013; revisado el 20 de abril de 2014; aceptado el 8 de septiembre de 2014.