

AUTORREFERENCIA, AUTOCONCIENCIA Y LA PARADOJA DE MOORE

JAVIER VIDAL

Departamento de Filosofía
Universidad de Concepción (Chile)
fravidal@udec.cl

RESUMEN: La paradoja de Moore es el tipo de absurdo de, entre otras cosas, la aserción de una oración de la forma “ p pero yo no creo que p ”. Propongo una explicación de la paradoja en términos del uso del pronombre personal “yo”. Más aún, recorro a un modelo del uso competente al que me referiré como modelo de la autorreferencia autoconsciente. A continuación, muestro que la explicación introducida es independiente de otras explicaciones recientes basadas en la naturaleza consciente de la aserción y de la creencia. Pero, además, concluyo que está justificado pensar que ni la aserción ni la creencia son de naturaleza consciente.

PALABRAS CLAVE: primera persona, conciencia, aserción, creencia, Rosenthal, Shoemaker

SUMMARY: Moore’s paradox is that kind of absurdity of, among other things, the assertion of a sentence having the form of “ p but I do not believe that p ”. I put forward an explanation of the paradox in terms of the use of the personal pronoun “I”. Moreover, I resort to a model of the competent use to which I will refer as a model of the self-conscious self-reference. Then, I show that the explanation set forth is independent of other recent explanations based on the conscious nature of both the assertion and the belief. But, besides, I conclude that there is good reason for thinking that neither the assertion nor the belief are of a conscious nature.

KEY WORDS: first person, consciousness, assertion, belief, Rosenthal, Shoemaker

I

La paradoja de Moore es el tipo de absurdo, sinsentido o irracionalidad propio de, entre otras cosas, la aserción de una oración de la forma “ p pero yo no creo que p ”.¹ El carácter paradójico radica en que ambos miembros de la conjunción podrían ser verdaderos a la vez sin

¹ G.E. Moore incluyó en el mismo tipo de absurdo la aserción de una oración de la forma “Yo creo que p pero *no* p ” (Moore 1944, p. 204). En este caso, el carácter paradójico radica también en que ambos miembros de la conjunción podrían ser verdaderos a la vez sin que hubiera una contradicción (de hecho lo son siempre que un hablante tiene una creencia falsa) y, con todo, parece intuitivamente que la aserción es por igual absurda. Por razones de espacio, me voy a ocupar extensamente sólo del primer caso. No obstante, creo que una explicación basada en la naturaleza de la primera persona puede responder a ambos casos. Justificaré brevemente esta afirmación más adelante.

que hubiera una contradicción (es más, lo son siempre que un hablante ignora un hecho) y, sin embargo, parece intuitivamente que hay algo absurdo o irracional en realizar esta aserción mooreana.² Esto sugiere que el problema no está en el contenido proposicional de la aserción sino en la aserción misma del contenido proposicional. A continuación, propongo una explicación de la paradoja de Moore en términos del uso del pronombre personal “yo”, en la medida en que una aserción mooreana incluye, al menos habitualmente, un uso de “yo”. Recorro para ello a un modelo del uso competente de “yo”, al que me referiré como modelo de la autorreferencia autoconsciente, según el cual el usuario de “yo” sabe que está hablando de él mismo. Pero mi propuesta tiene un alcance mayor: muestro que una explicación basada en la naturaleza de la primera persona es independiente de otras explicaciones basadas en la naturaleza consciente de la aserción y en la naturaleza consciente de la creencia. Como resultado de mi análisis expongo razones para pensar que, por una parte, la aserción no es de naturaleza consciente y que, por otra parte, no existe una relación constitutiva entre creencia y conciencia.

Estoy entendiendo que una explicación basada en la naturaleza de la primera persona no es solamente, en sentido amplio, una explicación basada en el tipo de conciencia (o conocimiento) que alguien tiene, por ejemplo, de su creencia o de su aserción, pues las otras explicaciones a las que me he referido también consideran este hecho como parte sustancial de la explicación de la paradoja de Moore. La naturaleza de la primera persona tiene que ver, en sentido estrecho, con el tipo de conciencia (o conocimiento) que alguien tiene de que *él mismo* cree algo o asevera algo. Por eso, ésta es una explicación en términos del uso de “yo” en la aserción de una oración mooreana.

II

Supongamos que Edipo realiza la aserción sincera de la oración “Llueve pero el hijo de Layo no cree que llueve”. No parece haber ningún tipo de sinsentido o irracionalidad aquí del tipo que es característico en la aserción de una oración mooreana. A pesar de que el referente de “el hijo de Layo” es Edipo y, por lo tanto, el usuario de la oración, Edipo puede no estar aseverando tanto que llueve como que *él mismo* no cree que llueve. Lo cierto es que Edipo puede ignorar la identidad que guarda con la persona referida por “el hijo de Layo”: en ese caso Edipo no es consciente de que está hablando de él mismo y por

² Hablaré indistintamente de aserción mooreana o de aserción de una oración (o conjunción) mooreana.

eso no está aseverando, entre otras cosas, que él mismo no cree que llueve. Éste es un primer indicio de que el sinsentido de una aserción mooreana tiene que ver con la naturaleza de la primera persona.

Según el modelo del uso competente de “yo” al que me referiré como modelo de la autorreferencia autoconsciente y cuya formulación se sustenta parcialmente en el trabajo de otros autores (Peacocke 1983, pp. 133–139 y 2008, pp. 78–92; O’Brien 2007, pp. 74–88), deben satisfacerse dos condiciones para que un uso de “yo” sea autoconsciente, en el sentido de que el usuario de “yo” sea consciente de estar hablando de él mismo:

- i) la condición pragmática de que el usuario de “yo”, como el usuario de cualquier otra palabra, sepa que él mismo es el hablante o usuario de la palabra;
- ii) la condición semántica de que la referencia de “yo” sea fijada por la regla de autorreferencia según la cual el referente de “yo” en cada contexto es el usuario de “yo” en el contexto y que, por lo tanto, el hablante use “yo” basándose en el conocimiento de esa regla.

Christopher Peacocke formula el modelo de la autorreferencia autoconsciente en los siguientes términos:

El emisor tiene una conciencia de la acción (lingüística) con el contenido:

(1L) Yo estoy emitiendo “Yo soy F”.

El emisor acepta como verídica esta conciencia de la acción y juzga (1L). Nuestro emisor también sabe

(2L) Cualquier emisión de “yo” refiere al autor de esa emisión.

Desde (1L) y (2L) nuestro emisor sabe

(3L) “yo” en esta emisión de “Yo soy F” refiere a mí.

De modo que el uso de “yo” de nuestro emisor es plenamente autoconsciente. (Peacocke 2008, p. 87)

En mi formulación del modelo emplearé la traducción al discurso indirecto del discurso directo del usuario. Entonces, un usuario X de “yo”, por ejemplo, en la aserción de la oración “Yo estoy levantando el brazo”, es consciente de que está hablando de él mismo como resultado de los siguientes conocimientos:

- i) X sabe que él mismo, un hablante de español, es el usuario de “yo” (condición pragmática).³
 - ii) X sabe que el referente de “yo” en cada contexto es el usuario de “yo” en el contexto (condición semántica).⁴
- Así que, en efecto:
- iii) X sabe que su uso de “yo” refiere a él mismo (por i y ii).⁵

Es importante señalar que la condición pragmática i) introduce un primer sentido en el que un hablante X es autoconsciente: X sabe que él mismo es el hablante. De ahí no se sigue, sin embargo, que X sepa que está hablando de él mismo. Ahora bien, si el conocimiento descrito en i) es el conocimiento idiosincrásico, no observacional,

³ Éste es el punto en que el modelo de la autorreferencia autoconsciente que estoy proponiendo difiere del modelo propuesto por Lucy O’Brien. Ella sostiene, entre sus críticas a la solución de Peacocke, que el conocimiento que X tiene de que está usando “yo” no es el conocimiento de que *él mismo* es el usuario, porque eso significa que, entonces, ese conocimiento incluye ya una *referencia* de primera persona a X. Pues, en ese caso, la capacidad de referirse de primera persona a uno mismo en el lenguaje tendría que explicarse mediante una capacidad previa de referirse de primera persona a uno mismo en el pensamiento, que, según el modelo, requeriría que X tuviera el conocimiento de que él mismo es el pensador, introduciendo una nueva referencia de primera persona a X que ahora no parece poder explicarse en términos de una nueva capacidad (O’Brien 2007, pp. 68–69). Ésta es una objeción importante que, como ella reconoce, podría solventarse si hubiera una explicación de la capacidad de referirse de primera persona a uno mismo en el pensamiento que no estuviera basada en el modelo de la autorreferencia autoconsciente (O’Brien 2007, p. 66). Reconozco la dificultad señalada por O’Brien y no estoy en condiciones de ofrecer una explicación de esa capacidad que no consista en decir simplemente que la capacidad en cuestión es *primitiva*. Pero como las explicaciones deben terminar en algún lugar, no veo por qué no pueda ser que efectivamente la capacidad de referirse de primera persona a uno mismo en el pensamiento sea una capacidad primitiva.

⁴ O’Brien, por ejemplo, habla expresamente del conocimiento de la regla de autorreferencia (O’Brien 2007, p. 84). Esto no significa, obviamente, que el conocimiento en cuestión sea tal que X tenga la capacidad de articular explícitamente la regla. Pero, desde luego, X debe tener el concepto de ser un usuario de “yo” y el concepto de referirse a algo, y además debe ser capaz de seguir la regla y no meramente de conformarse a ella (O’Brien 2007, p. 87). De modo que la capacidad atribuible al sujeto que sigue la regla de autorreferencia es una capacidad conceptual compleja.

⁵ Aunque en esta formulación el conocimiento que X tiene de que su uso de “yo” refiere a él mismo es representado como un conocimiento inferencial, no debe entenderse como un compromiso con la idea de que el usuario de “yo” tiene que pasar por un proceso psicológico de naturaleza deductiva (O’Brien 2007, p. 83). Sólo estoy proponiendo que iii) el ejercicio consciente de la capacidad de referirse de primera persona a uno mismo en el lenguaje requiere i) y ii), es decir, requiere el ejercicio consciente de otras dos capacidades independientes.

que un agente tiene de sus propias acciones,⁶ en este caso, de su acción lingüística, el uso de “yo” es inmune al error por un fallo de identificación⁷ *del usuario*: no es posible que X crea que él mismo es el usuario de “yo”, pero su creencia sea falsa sólo porque la persona que realmente está usando “yo” no es él mismo sino algún otro. El conocimiento de la regla de autorreferencia explica que, además, el usuario de “yo” sea consciente de que está hablando de él mismo. Pues el conocimiento de la regla es el conocimiento que marca toda la diferencia entre la autoconciencia poseída por el usuario de *cualquier* palabra, el conocimiento de que él mismo es el hablante, y la autoconciencia poseída por el usuario de “yo”, el conocimiento de que su uso de “yo” refiere a él mismo. El conocimiento de la regla también contribuye a explicar que el uso de “yo” sea inmune al error por un fallo de identificación *del referente*: no es posible que X crea que él mismo es el referente de su uso de “yo”, pero su creencia sea falsa sólo porque la persona que realmente es el referente de “yo” no es él mismo sino algún otro (Vidal 2011, pp. 74–78).

Al especificar tanto la autoconciencia descrita en i), consistente en que X sabe que él mismo es el hablante, como la autoconciencia descrita en iii), consistente en que X sabe que está hablando de

⁶ Un conocimiento que no está basado en la observación de sus acciones sino en el hecho mismo de realizarlas y que, por lo tanto, cuenta como un conocimiento no observacional. La idea básica del tipo de conocimiento o conciencia en cuestión es común tanto a Peacocke, quien habla de la conciencia de la acción (*action-awareness*), como a O'Brien, quien habla de la conciencia del agente (*agent's awareness*). Es curioso, sin embargo, que la semejanza terminológica oculta una diferencia profunda. Como se pone de manifiesto en la cita que incluyo en el texto, Peacocke sostiene que la conciencia de su propia acción que tiene alguien consiste en el conocimiento de que él mismo está realizando la acción (mental o física), por ejemplo, una acción lingüística. Esto queda claro por el uso de “yo” en la cláusula (1L). En cambio, O'Brien, como resultado de la crítica a Peacocke examinada en la nota 3, argumenta que la conciencia del agente consiste en el conocimiento que el agente tiene de que sus acciones son suyas, o de sus acciones como propias, que no es el conocimiento de que él mismo está realizando la acción (mental o física). Ahora bien, a pesar de los esfuerzos de explicación (O'Brien 2007, pp. 114–124; pp. 182–190), sigo sin ver cómo es posible que un agente sea consciente de su acción como *suya* sin ser consciente de que *él mismo* está realizando la acción. Así que me parece que O'Brien está tratando de poner el carro delante de los bueyes.

⁷ Ésta es una aplicación derivada de la noción de *inmunidad al error por un fallo de identificación*, que fue introducida inicialmente para dar cuenta de la autoconciencia *expresada* mediante un uso de “yo”: si el uso de “yo” está basado en el conocimiento no observacional de que alguien está levantando el brazo, no es posible que X realice la aserción de “Yo estoy levantando el brazo” pero su aserción sea errónea sólo porque la persona que está realmente levantando el brazo no es el referente de “yo”, es decir, no es él mismo (Shoemaker 1994, p. 82).

él mismo, he tenido que recurrir al uso del reflexivo indirecto “él mismo” (*he himself*), que traduce en el discurso indirecto el uso del pronombre personal “yo” en el discurso directo. Si X realiza la aserción sincera de “Yo estoy levantando el brazo”, es correcto decir:

C) X cree que él mismo está levantando el brazo.

Ahora bien, una atribución de creencia como C) se caracteriza porque no es posible que la persona referida por “X” (fuera de la cláusula-que), quien es el sujeto de la creencia, ignore la identidad que guarda con la persona referida por “él mismo” (dentro de la cláusula-que), quien es el objeto de la creencia.⁸ De modo que de C) se sigue:

CC) X sabe que la creencia de que alguien está levantando el brazo es sobre quien cree que alguien está levantando el brazo, es decir, él mismo.

Podemos, entonces, dar una explicación del sinsentido o irracionalidad que es propio de una aserción mooreana en términos del modelo de la autorreferencia autoconsciente.⁹ Si X realiza la aserción sincera de la oración mooreana “Llueve pero yo no creo que llueve”,

⁸ Una diferencia clave entre el reflexivo directo “él” (que también puede usarse en el discurso indirecto) y el reflexivo indirecto (o cuasiindicador) “él mismo” es, precisamente, que en una atribución de creencia como:

Cx) X cree que él está levantando el brazo

es posible que X ignore la identidad que guarda con la persona referida por “él”. En efecto, supongamos que X es la persona reflejada en un espejo deformante, pero X no se reconoce en ella. Si X realiza la aserción sincera de “La persona reflejada en el espejo está levantando el brazo”, es correcto decir en el discurso indirecto algo de la forma Cx): es una propiedad de “él” poder reemplazar *salva veritate* (en el discurso indirecto) cualquier término singular o descripción definida que el hablante X haya usado. Sin embargo, X ignora la identidad que guarda con la persona referida por “él” (Bermúdez 1998, pp. 2–4; Castañeda 1999).

⁹ Uno de los árbitros del artículo ha planteado la siguiente objeción a una explicación de la paradoja de Moore basada en un modelo del uso competente de “yo”: que en español la aserción de una oración mooreana como “Llueve pero no creo que llueve” no incluye un uso del pronombre personal y, sin embargo, no por ello desaparece la paradoja. Una respuesta simple sería argumentar que en español un uso de “Creo”, como un uso de “No creo”, es derivado y que la forma verbal de primera persona incorpora, como una función adicional, la función referencial que es propia del pronombre personal. De manera que la condición semántica podría formularse así: X sabe que el referente *adicional* de “Creo” en cada contexto es el usuario de “Creo” en el contexto. Entonces, las condiciones para que haya una autorreferencia autoconsciente pueden formularse así:

resulta que podemos inferir el sinsentido o irracionalidad de hacerlo a partir de los siguientes conocimientos y creencias:

- 1) X sabe que él mismo, un hablante de español, está aseverando sinceramente la oración “Llueve pero yo no creo que llueve” (condición pragmática).
- 2) X sabe que el referente de “yo” en cada contexto es el usuario de “yo” en el contexto (condición semántica).
- 3) X sabe que su uso de “yo” en la aserción de la oración “Llueve pero yo no creo que llueve” refiere a él mismo (por 1 y 2).
- 4) X sabe que él mismo está aseverando sinceramente que llueve y que él mismo no cree que llueve (por 1 y 3).

Al establecer en 1) que X es un hablante competente, es razonable sostener que él sabe que una aserción sincera con el contenido proposicional p expresa una creencia con el contenido proposicional p . Pues, X realiza intencionalmente una aserción sincera para informar de lo que cree. Desde luego, el término “creencia” no está siendo usado aquí en un sentido disposicional sino para referirse a una creencia, o un pensamiento, como un tipo de suceso que está ocurriendo en la mente del hablante a la vez que la aserción es realizada (Rosenthal 2002, p. 170). De manera que de 4) puede inferirse:

- 5) X sabe que él mismo cree (que llueve y que él mismo no cree que llueve).

Como el verbo “saber” es un verbo factivo, en el sentido de que si X sabe que p entonces p (Williamson 2000, pp. 34–39), y la persona referida por “X” es la persona referida por “él mismo”, obtenemos:

- 6) X cree (que llueve y que él mismo no cree que llueve).

i) X sabe que él mismo, un hablante de español, es el usuario de “Creo” (condición pragmática).

ii) X sabe que el referente adicional de “Creo” en cada contexto es el usuario de “Creo” en el contexto (condición semántica).

Así que, en efecto:

iii) X sabe que su uso de “Creo” refiere adicionalmente a él mismo (por i y ii).

Pero resulta que, como es propio del reflexivo indirecto, la persona referida por “X” (fuera de la cláusula-que), quien cree que llueve y que alguien no cree que llueve, no puede ignorar la identidad que guarda con la persona referida por “él mismo” (dentro de la cláusula-que), *de* quien X cree que no cree que llueve. Esto último se sigue trivialmente del hecho de que la creencia se distribuye sobre cada miembro de 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 llueve y X cree que él mismo no cree que llueve. Entonces, de 6) 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 no cree que llueve es sobre quien cree que llueve, es decir, él mismo.¹⁰

En otras palabras: X sabe que la persona de quien cree que no cree que llueve es la *misma* persona, él mismo, que cree que llueve. La relevancia explicativa del uso de “yo” está en lo siguiente: del hecho de que X crea que él mismo no cree que llueve, que es la creencia expresada por la segunda parte de la conjunción mediante un uso de “yo”, se sigue que X sabe que la creencia de que alguien no cree que llueve es sobre quien cree que llueve (es decir, él mismo), que es la creencia expresada por la primera parte de la conjunción. Es seguro que alguien en posesión del conocimiento descrito en 7) no puede seguir creyendo racionalmente tanto que llueve como que él mismo no cree que llueve. Tampoco es racional, por ende, que realice la aserción sincera de la oración mooreana “Llueve pero yo no creo que llueve”.

Como señalé en la nota 1, voy a explicar brevemente en qué consistiría la irracionalidad de una aserción de la oración mooreana “Yo creo que llueve pero no llueve”. Resulta que, según el modelo de la autorreferencia autoconsciente, llegaríamos a inferir, análogamente a 6):

- 8) X cree (que él mismo cree que llueve y que no llueve).

¹⁰ La versión completa desde 6) es:

7*) X sabe que la creencia (de que llueve y de que alguien no cree que llueve) es sobre quien cree (que llueve y que alguien no cree que llueve), es decir, él mismo.

Del mismo modo que CC) se sigue de C). Véase antes.

Pero de 8) obtenemos, análogamente a 7):

- 9) X sabe que la creencia de que alguien cree que llueve es sobre quien cree que no llueve, es decir, *él mismo*.

Así que la relevancia explicativa del uso de “yo” está en lo siguiente: del hecho de que X crea que *él mismo* cree que llueve, que es la creencia expresada por la primera parte de la conjunción mediante un uso de “yo”, se sigue que X sabe que la creencia de que alguien cree que llueve es sobre quien cree que no llueve, que es la creencia expresada por la segunda parte de la conjunción. Desde luego, es posible que alguien con la mente dividida tenga tanto la creencia de que llueve como la creencia de que no llueve. Pero carece de sentido que X sepa que la creencia de que alguien cree que llueve es sobre quien cree que no llueve, es decir, *él mismo*. Pues entonces lo racional sería abandonar una de esas creencias o, más bien, ambas. Tampoco es racional, en consecuencia, que X realice la aserción de la oración mooreana “Yo creo que llueve pero no llueve”. De manera que una explicación basada en la naturaleza de la primera persona puede ser una explicación satisfactoria de la irracionalidad o sinsentido de cualquier aserción mooreana.¹¹

III

Esta explicación es independiente de otras explicaciones que, sin negar la importancia del uso de “yo” en una aserción mooreana, recurren a consideraciones sobre la naturaleza consciente de la aserción o la naturaleza consciente de la creencia. Por ejemplo, una explicación puede basarse en la tesis de David Rosenthal (1995, pp. 316–320) de que cualquier aserción sincera expresa una creencia consciente,¹² incluyendo una aserción sincera que no es la autoadscripción de una creencia, como la aserción sincera de “Llueve”. La idea es que la creencia de que llueve que un hablante expresa mediante la

¹¹ Pero debido a la extensión del artículo, en la comparación con otras explicaciones de la paradoja de Moore sólo voy a considerar la aserción mooreana de “Llueve pero yo no creo que llueve”. Además, esta limitación está justificada porque ya no se trata de probar que la explicación basada en la naturaleza de la primera persona funciona en todos los casos.

¹² En realidad, Rosenthal ofrece otra explicación de la paradoja de Moore que no está basada en la naturaleza consciente de la aserción sino en la distinción entre las condiciones de verdad y las condiciones de uso de la aserción mooreana. Pero sostiene que cualquier estado intencional verbalmente expresado es consciente y, en particular, que la creencia expresada por una aserción sincera es consciente (Rosenthal 2010, pp. 32–33).

aserción sincera de “Llueve” es regularmente, cuando alguien realiza esa aserción, la causa de una creencia de segundo orden cuyo contenido es que él mismo cree que llueve, que es en lo que consiste que la creencia expresada sea consciente. Ello es así porque las oraciones “Llueve” y “Yo creo que llueve” tienen las mismas condiciones de uso (o son performativamente equivalentes) y, por tanto, quien realiza la aserción sincera de “Llueve” tendrá la creencia de segundo orden que puede ser expresada mediante la aserción sincera de “Yo creo que llueve” (Rosenthal 2010, pp. 27–29). Sucede, entonces, que quien realiza la aserción sincera de la oración mooreana “Llueve pero yo no creo que llueve” tiene una creencia de segundo orden sobre la creencia de primer orden, la creencia de que llueve, que es expresada por la primera parte de la conjunción. De manera que el hablante, al tener la creencia de que él mismo cree que llueve, estará dispuesto a aseverar sinceramente la oración “Yo creo que llueve”, que es una autoadscripción de la creencia de que llueve, porque al ser consciente de la creencia de que llueve él está en condiciones de autoadscribirla. Pero, de hecho, el hablante está aseverando sinceramente la negación de esa oración mediante la segunda parte de la conjunción mooreana. El resultado es que el hablante está irracionalmente dispuesto a realizar tanto la aserción de “Yo creo que llueve” como la aserción de “Yo no creo que llueve”, que efectivamente realiza como parte de la aserción mooreana.

Es cierto que la explicación presentada en la sección anterior incluye la idea de que, al satisfacerse las condiciones descritas en el modelo de la autorreferencia autoconsciente, la aserción es de naturaleza consciente: como resultado, en último caso, de la condición pragmática, es decir, de la autoconciencia consistente en el conocimiento de que uno mismo está realizando sinceramente la aserción mooreana, pudo derivarse 5), y por distribución de la creencia sobre el primer miembro de la conjunción obtenemos:

10) X sabe que él mismo cree que llueve.

De modo que la primera parte de la conjunción, o equivalentemente la aserción sincera de “Llueve”, expresa la creencia consciente de que llueve.¹³ Desde luego, esto no significa comprometerse con la tesis de que el conocimiento descrito en 10) involucra una creencia de segundo orden sobre una creencia de primer orden. Como señalé en la

¹³ Pero defenderé más adelante que la aserción no es de naturaleza consciente porque es concebible que alguien realice una aserción sin que la condición pragmática sea satisfecha. Véase la sección IV.

nota 4, me parece correcta la tesis de O'Brien de que el conocimiento de la regla de autorreferencia descrito en 2) es el ejercicio consciente de una capacidad conceptual compleja. Así que, por derivación de todos los pasos anteriores, resulta que el conocimiento descrito en 10) tiene que ser un tipo de conciencia conceptual. Pero no veo por qué si alguien tiene un conocimiento que es el ejercicio consciente de una capacidad conceptual entonces está en posesión de una creencia, dado que el conocimiento no es analizable en términos de creencia (Williamson 2000, pp. 27–33) y, como consecuencia, no veo por qué si alguien tiene el conocimiento de que él mismo cree que llueve, que es el ejercicio consciente de una capacidad conceptual, entonces está en posesión de la creencia (de segundo orden) de que él mismo cree que llueve. Con todo, el fenómeno es el mismo: haya o no haya una creencia de segundo orden, el hecho es que una aserción sincera expresa una creencia consciente.

Ahora bien, supongamos que el hablante X, que *es* la persona o el cuerpo humano vivo que cae bajo su observación especial, el tipo de observación que cada uno tiene de sí mismo habitualmente, usa el término “A” para referirse a la persona que cae bajo su observación especial en cada contexto. Fue G.E.M. Anscombe quien introdujo el caso imaginario del usuario de un cierto término singular “A”:

Imaginemos una sociedad en la que todos están etiquetados con dos nombres. Uno figura en su espalda y en la parte superior de su pecho, y estos nombres, que sus portadores no pueden ver, son varios: de la “B” a la “Z”, por decir algo. El otro, “A”, está estampado en el interior de sus muñecas, y es el mismo para todos. Al hacer informes sobre las acciones de los demás, todos usan el nombre sobre su pecho o su espalda si pueden ver estos nombres o están acostumbrados a verlos. Además, todos aprenden a responder a emisiones del nombre sobre su propio pecho y espalda en el modo y en el tipo de circunstancias en las que nosotros tendemos a responder a emisiones de nuestros nombres. Informes sobre las propias acciones de uno, que uno da a partir de la observación, se hacen usando el nombre sobre la muñeca. (Anscombe 1981, p. 24)

Parece que “A” es un término singular que, como el pronombre personal “yo”, garantiza la referencia al hablante. Teniendo en cuenta que el usuario de “A” observa desde un punto de vista especial su propia muñeca, la referencia de “A” puede ser fijada mediante la siguiente regla: el referente de “A” en cada contexto es la persona que cae bajo la observación especial del usuario de “A” en el contexto.

Pero, en condiciones normales, la persona que cae bajo la observación especial del usuario de “A” es precisamente el usuario de “A”, es decir, el hablante:

De este modo, suponiendo que conozco a alguien llamado “X” y llamo a una cosa “X” con la intención de referirme a esa persona, una garantía de referencia en este sentido sería una garantía de que hay tal cosa como X. El nombre “A” que inventé tendría este tipo de referencia garantizada. El usuario de “A” quiere hablar de cierto ser humano, alguien que cae bajo su observación de un modo más bien especial.¹⁴ Esa persona es él mismo y, por ello, dado que él ha captado el uso de “A”, él no puede menos que estar hablando de una persona real. (Anscombe 1981, p. 30)

Imaginemos ahora que X realiza la aserción sincera de la oración “Llueve pero A no cree que llueve”, pudiendo inferir los siguientes conocimientos y creencias:¹⁵

- 11) X sabe que él mismo, un hablante de español, está aseverando sinceramente la oración “Llueve pero A no cree que llueve” (condición pragmática).
- 12) X sabe que el referente de “A” en cada contexto es la persona que cae bajo la observación especial del usuario de “A” en el contexto.
- 13) X sabe que su uso de “A” refiere a la persona que cae bajo su observación especial (por 11 y 12).
- 14) X sabe que él mismo está aseverando sinceramente que llueve y que la persona que cae bajo su observación especial no cree que llueve (por 11 y 13).

¹⁴ La noción de una observación *especial* se glosa de la siguiente manera:

Así, para cada persona hay una persona de quien tiene perspectivas tanto característicamente limitadas como característicamente privilegiadas: de no ser en espejos, ella nunca ve a la persona entera, y sólo puede lograr perspectivas más bien especiales de lo que ve. Algunas son particularmente buenas y otras son particularmente malas. (Anscombe 1981, p. 24)

¹⁵ Para la justificación de la transición de unos conocimientos a otros, y evitar repeticiones, me remito a la justificación dada en la sección II. También en las próximas inferencias me limitaré a dar una justificación de las transiciones nuevas.

- 15) X sabe que él mismo cree (que llueve y que la persona que cae bajo su observación especial no cree que llueve).
- 16) X cree (que llueve y que la persona que cae bajo su observación especial no cree que llueve).

Pero resulta que la persona referida por “X”, quien cree que llueve y que alguien no cree que llueve, puede ignorar la identidad que guarda con la persona referida por “la persona que cae bajo su observación especial”, de quien X cree que no cree que llueve. Pues, a pesar de que el referente de “A”, la persona que cae bajo la observación especial del usuario de “A”, es el propio usuario de “A”, no es posible inferir de 11) y de 12):

- 17) X sabe que su uso de “A” refiere a él mismo.¹⁶

Sería necesario que, contrariamente a 12), el conocimiento semántico respecto de “A” fuese el conocimiento de la regla de autorreferencia:

- 18) X sabe que el referente de “A” en cada contexto es el usuario de “A” en el contexto (condición semántica).

Es obvio, entonces, que de 16) y de la distribución de la creencia sobre cada miembro de la conjunción *no* se sigue:

- 19) X sabe que la creencia de que alguien no cree que llueve es sobre quien cree que llueve, es decir, él mismo.

No hay, por tanto, ningún tipo de sinsentido o irracionalidad en la aserción sincera de la oración “Llueve pero A no cree que llueve”. Como era de esperar, teniendo en cuenta que no es una aserción mooreana con un uso del pronombre personal “yo”. Sin embargo, la autoconciencia característica de la condición pragmática garantiza la verdad de 15), y por distribución de la creencia sobre el primer miembro de la conjunción obtenemos:

¹⁶ Por ello, así como Edipo puede usar la descripción “el asesino de Layo” para referirse a él mismo sin ser consciente de ello, los usuarios de “A” se refieren a ellos mismos sin ser conscientes de que lo hacen. Otra forma de subrayar el mismo punto es decir que el hablante X usa “A” para dar informes de sus propias acciones pero, como Anscombe señala, eso sólo significa que son informes sobre las acciones de X que proceden de la boca de X (Anscombe 1981, p. 24). Si X realiza la aserción de “A está levantando el brazo”, estamos ante un informe de una acción de X que procede de la boca de X. Está claro que, al realizar la aserción, X expresa la creencia de que la persona que cae bajo su observación especial está levantando el brazo. Esa persona es precisamente X. Sin embargo, mediante su uso de “A”, X no expresa la creencia de que *él mismo* está levantando el brazo.

20) X sabe que él mismo cree que llueve.

Esto significa que la tesis de que una aserción sincera expresa una creencia consciente, como mostraría 20), es insuficientemente explicativa del sinsentido o irracionalidad que está en juego en la aserción de una oración mooreana. Concluyo, entonces, que la explicación de la paradoja de Moore basada en la naturaleza de la primera persona tiene prioridad sobre la explicación basada en la naturaleza consciente de la aserción.

IV

Señalé anteriormente que un uso de “yo” es inmune al error por un fallo de identificación del usuario *si* el conocimiento del usuario es el conocimiento idiosincrásico, no observacional, que un agente tiene de su propia acción lingüística. Pero es posible que el conocimiento del usuario sea un conocimiento observacional de su propia aserción.¹⁷ Pues bien, supongamos, a la manera de Anscombe, que existe una sociedad imaginaria en la que los hablantes sólo logran referirse a sí mismos mediante el término “A”. Supongamos ahora, a diferencia de Anscombe, que posteriormente los hablantes son adiestrados en el uso de “yo” y, por consiguiente, en el conocimiento de la regla de autorreferencia. Pero aprenden a usar “yo” como reemplazo de “A”: aprenden a usar “yo” para reportar sus acciones y sus estados mentales (como sus creencias) a partir de la observación de las acciones de la persona que cae bajo su observación especial, porque parte de su adiestramiento pasa por instruirles con la información adicional de que el referente de “A” es el referente de “yo”.

Ahora podemos construir un caso en que el hablante de una aserción mooreana no sabe que él mismo es el hablante, de manera que reportará su acción lingüística (y, por ello, sus creencias) a partir de la observación de que la persona que cae bajo su observación especial realiza la aserción mooreana:

¹⁷ O’Brien discute este punto y sostiene la posibilidad de que un hablante use ocasionalmente “yo” de un modo competente teniendo sólo un conocimiento observacional de la propia acción lingüística. En ese contexto, introduce una situación como la imaginada por David Lewis en la que hay dos dioses que son omniscientes, lo que incluye el conocimiento del dios que está usando “yo”, pero de tal modo que ninguno de ellos sabe cuál de los dos dioses es él mismo. Entonces, es posible que, por ejemplo, el dios de la montaña más alta use “yo” teniendo el conocimiento de que el dios de la montaña más alta es el usuario de “yo”, pero sin saber que él mismo es el dios de la montaña más alta: en esa situación, el dios de la montaña más alta no sabe que él mismo es el hablante (O’Brien 2007, p. 58, pp. 84–86).

- 21) X sabe que la persona que cae bajo su observación especial, un hablante de español, está aseverando la oración “Llueve pero yo no creo que llueve”.

Está claro que X no puede tener un conocimiento observacional de que la aserción es *sincera* aunque, desde luego, supondrá por defecto que lo es. Pero no se trata de algo que se sepa.

- 22) X sabe que el referente de “yo” en cada contexto es el usuario de “yo” en el contexto (condición semántica).
- 23) X sabe que el uso de “yo” en la aserción de la oración “Llueve pero yo no creo que llueve” refiere a la persona que cae bajo su observación especial (por 21 y 22).
- 24) X sabe que la persona que cae bajo su observación especial está aseverando que llueve y que ella misma no cree que llueve (por 21 y 23).

Al establecer en 21) que X, es decir, la persona que cae bajo su observación especial, es un hablante competente, es razonable sostener que él sabe que mediante una aserción con el contenido proposicional *p* un hablante se comporta, ante un posible oyente, como si creyera el contenido proposicional *p*, aunque pueda no ser así, como ocurre paradigmáticamente cuando alguien miente. Me referiré a este comportamiento lingüístico como expresión de una creencia*. Este comportamiento puede entenderse simplemente como el hecho consistente en que, al realizar la aserción de una oración, el hablante se comporta como si tuviera la creencia que expresaría mediante una aserción sincera de esa oración. Decir de alguien que cree* algo no es más que decir que se comporta como si lo creyera, esté o no esté realizando una cierta aserción, esto es, esté o no esté expresando esa creencia*, pues el comportamiento en cuestión puede no ser lingüístico. Pero está claro que cualquier aserción, incluyendo una aserción sincera, expresa una creencia* en el sentido indicado.¹⁸ De manera que de 24) puede inferirse:

¹⁸ J.N. Williams habla de expresar ostensiblemente la creencia: una aserción sincera expresa una creencia y una aserción, sea o no sea sincera, expresa *ostensiblemente* una creencia. Mi aserción de que *p* expresa ostensiblemente la creencia de que *p* sólo si yo me comporto de tal forma que te doy intencionalmente una razón para que pienses que tengo esa creencia. Pero puedo no tener realmente la creencia de que *p* (Williams 2006, p. 241). Es, por tanto, aproximadamente lo mismo decir que una aserción expresa ostensiblemente una creencia y decir que una aserción expresa una creencia*. He decidido introducir esta fórmula para no hacer ciertos cambios que impedirían una lectura más amable del texto.

- 25) X sabe que la persona que cae bajo su observación especial cree* (que llueve y que ella misma no cree que llueve).

Como el verbo “saber” es un verbo factivo y X es precisamente la persona que cae bajo su observación especial, resulta finalmente:

- 26) X cree* (que llueve y que él mismo no cree que llueve).

Como el reflexivo indirecto está en juego, no es posible que la persona referida por “X”, quien cree* que llueve y que alguien no cree que llueve, ignore la identidad que guarda con la persona referida por “él mismo”, de quien X cree* que no cree que llueve. De 26) y de la distribución de la creencia* sobre cada miembro de la conjunción se sigue:

- 27) X sabe que la creencia* de que alguien no cree que llueve es sobre quien cree* que llueve, es decir, él mismo.¹⁹

Alguien con el conocimiento descrito en 27) no puede comportarse racionalmente al realizar la aserción mooreana: sabiendo que, al realizar la aserción, se comporta como si creyera tanto que llueve como que él mismo no cree que llueve, esa aserción no podría constituir un comportamiento racional. Desde el punto de vista de las intenciones comunicativas necesarias para que una aserción sea exitosa, en el sentido de que el hablante pueda producir el efecto deseado en la audiencia a través de su comportamiento lingüístico, no es racional que X realice la aserción de la oración mooreana “Llueve pero yo no creo que llueve”.²⁰

Por otra parte, a falta de que X tenga la autoconciencia consistente en el conocimiento de que él mismo es el hablante que está realizando

¹⁹ Esto es,

27*) X sabe que comportarse como si creyera que alguien no cree que llueve es un comportamiento (como de creyente) sobre quien se comporta como si creyera que llueve, es decir, él mismo.

La distribución de la creencia* sobre la conjunción parece, también, algo evidente: si X se comporta como si creyera que *p* y *q*, entonces X se comporta como si creyera que *p* y X se comporta como si creyera que *q*.

²⁰ Pues el hablante sabe que, a partir de la aserción mooreana, no puede ser satisfecha la intención o bien de que la audiencia crea que él tiene una determinada creencia o bien de que la audiencia crea que él no tiene una determinada creencia. En consecuencia, tampoco puede intentar o bien que la audiencia crea que llueve o bien que la audiencia deje de creer que llueve. Se sigue de esto que, como ha sido enfatizado por diversos autores, tampoco un hablante racional podría realizar una aserción mooreana con la intención de mentir (Williams 2006, p. 229).

la aserción mooreana, sólo podemos obtener de 25) por distribución de la creencia* sobre el primer miembro de la conjunción:

- 28) X sabe que la persona que cae bajo su observación especial cree* que llueve.

Esto es, a falta de que la condición pragmática sea satisfecha, no podemos inferir de 25) que la primera parte de la conjunción expresa la creencia* *consciente* de que llueve, pues decir que una aserción expresa la creencia* consciente de que llueve es decir que el hablante sabe que, al realizar la aserción, *él mismo* (sepa o no sepa que es la persona que cae bajo su observación especial) se comporta como si creyera que llueve. En términos de Rosenthal, es decir que el hablante cree que, al realizar la aserción, él mismo se comporta como si creyera que llueve.

Pero teniendo en cuenta la verdad de 27), es obvio:

- 29) X sabe que él mismo cree* que llueve.

El resultado es que, a pesar de que X no tiene el conocimiento de que él mismo es el hablante, su aserción de la oración mooreana expresa la creencia* consciente de que llueve. Pero el carácter consciente de la creencia* no fue inferido, en este caso, del hecho de que la condición pragmática fuera satisfecha.

Argumento a continuación que el análisis anterior nos da una razón a favor de que la aserción no es de naturaleza consciente. En consecuencia, no podemos recurrir a la naturaleza consciente de la aserción para explicar la paradoja de Moore. En efecto, supongamos que un hablante X de la sociedad imaginaria realiza la aserción de “Llueve”, de lo que podemos inferir:

- 30) X sabe que la persona que cae bajo su observación especial, un hablante de español, está aseverando la oración “Llueve”.²¹

²¹ Rosenthal podría argumentar, como uno de los árbitros de este artículo me lo ha señalado, que una condición necesaria de cualquier aserción es que el hablante tenga un conocimiento no observacional de que él mismo está realizando la aserción, en cuyo caso el presunto hablante de la sociedad imaginaria no estaría en condiciones de realizar aserción alguna y, por lo tanto, no estaríamos ante un contraejemplo de la tesis de que la aserción es de naturaleza consciente. Pero me parece que Rosenthal no puede responder simplemente de esta manera. Puesto que precisamente uno de los méritos de la sociedad imaginaria de Anscombe, cuyo propósito central, como es sabido, es defender la tesis de que “yo” no es un término referencial, es habernos dado algunas razones para pensar que un hablante podría carecer del conocimiento no observacional de su propia aserción. Es decir, parece que Anscombe nos ha persuadido de que esto es algo concebible.

31) X sabe que la persona que cae bajo su observación especial está aseverando que llueve.

32) X sabe que la persona que cae bajo su observación especial cree* que llueve.

Así que, en efecto:

33) X cree* que llueve.

La aserción expresa la creencia* que X tiene de que llueve pero no hay ninguna razón para concluir que X sabe (o cree) que él mismo cree* que llueve. Pues, a falta de la autoconciencia consistente en el conocimiento de que él mismo es el hablante que está realizando la aserción, no parece haber nada más que justifique la conclusión de que X sabe que la creencia* de que llueve es *su* creencia*. Esto significaría que la aserción *no* es de naturaleza consciente, en contra de la tesis sostenida por Rosenthal: en este caso, la aserción no expresa la creencia* consciente de que llueve y, por tanto, no toda aserción expresa una creencia* consciente.²² Pero ésta es una razón para concluir que la explicación de la paradoja de Moore en términos de la naturaleza consciente de la aserción finalmente debe ser desestimada. Desde luego, quien realiza la aserción sincera de la oración mooreana “Llueve pero yo no creo que llueve” tendrá una creencia de segundo

²² Entiendo, pues, que la tesis de que la aserción es de naturaleza consciente no debe aplicarse sólo a las aserciones sinceras: si la aserción de “Llueve” expresa la creencia* de que llueve, la tesis debiera entrañar que el hablante tiene la creencia de que él mismo se comporta como si creyera que llueve. De ahí que si la tesis de Rosenthal fuese correcta, la explicación de la paradoja de Moore en términos de la naturaleza consciente de la aserción funcionaría igualmente. El hablante, al tener una creencia sobre la creencia* expresada en la primera parte de la conjunción, estaría dispuesto a realizar la aserción sincera de “Me comporto como si creyera que llueve”. La segunda parte de la conjunción reporta la falta de una creencia y, como hemos visto, expresa la creencia* de que él mismo no cree que llueve (o, equivalentemente, el hablante se comporta como si creyera que él mismo no cree que llueve). Sin embargo, parece indudable que por el hecho mismo de reportar esa falta de creencia el hablante también expresa la creencia* consistente en comportarse como si no creyera que llueve. (Del mismo modo, quien realiza la aserción de “Yo creo que llueve” expresa tanto la creencia* de que él mismo cree que llueve como, por el hecho mismo de reportar su creencia, la creencia* de que llueve). Así que, por la tesis de Rosenthal, tendrá la creencia de que él mismo se comporta como si no creyera que llueve y, por tanto, también estará dispuesto a realizar la aserción sincera de “Me comporto como si no creyera que llueve”. Luego, el hablante estará irracionalmente dispuesto a realizar tanto la aserción sincera de “Me comporto como si creyera que llueve” como la aserción sincera de “Me comporto como si no creyera que llueve”.

orden, o algún tipo de conocimiento, sobre la creencia de primer orden expresada por la primera parte de la conjunción. Entonces estará dispuesto a aseverar sinceramente la oración “Yo creo que llueve”, que es una autoadscripción de la creencia de que llueve, porque al ser consciente de esta creencia estará en condiciones de autoadscribirse. Sin embargo, de hecho él está aseverando sinceramente la negación de esa oración mediante la segunda parte de la conjunción mooreana. Pero la explicación no está en que, al realizar la aserción mooreana, la creencia expresada por la primera parte de la conjunción sea la causa de una creencia de segundo orden porque la creencia expresada por cualquier aserción (o parte de una aserción) que no es la autoadscripción de una creencia es regularmente la causa de una creencia de segundo orden. La explicación está en que la aserción mooreana incluye un uso del pronombre personal “yo” y, por ello, el hablante tiene una creencia de segundo orden, o algún tipo de conocimiento, sobre la creencia de que llueve.

V

Voy a defender ahora que la explicación del tipo de sinsentido o irracionalidad mooreana en términos de la naturaleza de la primera persona es también independiente de otra explicación basada en la naturaleza consciente de la creencia. Es Sydney Shoemaker (1996, pp. 76–83; 2009, pp. 37–43) quien sostiene que existe una relación constitutiva entre una creencia de primer orden y una creencia de segundo orden sobre ella. Esta relación se considera constitutiva porque no es causal, como en la propuesta de Rosenthal, sino racional en este sentido: es una exigencia racional que si la creencia de primer orden está de algún modo disponible (es, digamos, accesible), entonces el creyente tendrá una creencia de segundo orden sobre ella. Ello se debe, entre otras cosas, a que el razonamiento y la conducta de alguien que sólo tiene una creencia de primer orden que está disponible y de alguien que, además, tiene la creencia de segundo orden son indiscernibles (Shoemaker 1996, pp. 82–83). Y si la creencia de primer orden está de algún modo disponible y el creyente se plantea la cuestión de su verdad, sería irracional que no asintiera tanto mental como lingüísticamente al contenido de la creencia (Shoemaker 1996, pp. 80–81). De manera que, para el hablante que realiza una aserción sincera de “Llueve”, la creencia de que llueve está disponible y, por ende, tiene la creencia de que él mismo cree que llueve. Pero no es, como sostiene Rosenthal, que la aserción sincera de “Llueve” sea de naturaleza consciente sino que el hecho de que la creencia de que

llueve esté disponible, que es una condición necesaria de esa aserción sincera, entraña constitutivamente que el creyente tenga la creencia de que él mismo cree que llueve, que es en lo que consiste que la creencia de que llueve sea consciente.²³ Así que el hablante que realiza la aserción sincera de la oración “Llueve pero yo no creo que llueve” y, por tanto, cree que llueve, que es la creencia expresada por la primera parte de la conjunción, tendrá la creencia de que él mismo cree que llueve. Pero, entonces, la creencia de que él mismo cree que llueve también estará disponible, en cuyo caso sería irracional que, al plantearse la cuestión de su verdad, el creyente no estuviera dispuesto a asentir tanto mental como lingüísticamente al contenido de esa creencia. En consecuencia, estará dispuesto a realizar la aserción sincera de “Yo creo que llueve”, cuando de hecho está aseverando sinceramente su negación mediante la segunda parte de la conjunción mooreana. El resultado es que el hablante está irracionalmente dispuesto a realizar tanto la aserción de “Yo creo que llueve” como la aserción de “Yo no creo que llueve”, que efectivamente realiza como parte de la aserción mooreana. Entiendo que la misma explicación debería aplicarse a una aserción mooreana tal que la primera parte de la conjunción expresa la creencia* de que llueve. De no hacerlo, la explicación colapsaría.²⁴

²³ Pero, en su último artículo, Shoemaker sostiene que la disponibilidad y la conciencia de una creencia pueden ser consideradas como equivalentes, en el sentido de “acceso consciente” (o conciencia de acceso) a la creencia (Shoemaker 2009, p. 40). La naturaleza consciente de una creencia de primer orden no consiste, entonces, en tener una creencia de segundo orden sino en su disponibilidad. Pero, en cualquier caso, la disponibilidad de la creencia de primer orden entraña tener una creencia de segundo orden. Además, la propuesta reciente de Shoemaker es, más específicamente, que la relación constitutiva entre una creencia de primer orden y una creencia de segundo orden sobre ella consiste en que la creencia de segundo orden contiene como *parte* suya la creencia de primer orden (Shoemaker 2009, p. 42).

²⁴ Cuando, al realizar la aserción mooreana, alguien se comporta como si creyera que llueve y ese comportamiento está disponible, es decir, es accesible para él, la creencia* de que llueve será consciente en el sentido de que el hablante tendrá la creencia de que él mismo se comporta como si creyera que llueve. Por tanto, estará dispuesto a realizar la aserción sincera de “Me comporto como si creyera que llueve”. Pero mediante la segunda parte de la conjunción el hablante expresa la creencia* consistente en comportarse como si no creyera que llueve, que también estará disponible (Véase la nota 22). Así que, según Shoemaker, estará dispuesto a tener la creencia de que él mismo se comporta como si no creyera que llueve y, por ello, estará dispuesto a realizar la aserción sincera de “Me comporto como si no creyera que llueve”. Luego, el hablante estará irracionalmente dispuesto a realizar tanto la aserción sincera de “Me comporto como si creyera que llueve” como la aserción sincera de “Me comporto como si no creyera que llueve”.

Pues bien, supongamos que el hablante de la sociedad imaginaria realiza la aserción de la oración “Yo estoy levantando el brazo”, pudiendo inferir:

- 34) X sabe que la persona que cae bajo su observación especial, un hablante de español, está aseverando la oración “Yo estoy levantando el brazo”.²⁵
- 35) X sabe que el referente de “yo” en cada contexto es el usuario de “yo” en el contexto (condición semántica).
- 36) X sabe que el uso de “yo” en la aserción de la oración “Yo estoy levantando el brazo” refiere a la persona que cae bajo su observación especial (por 34 y 35).
- 37) X sabe que la persona que cae bajo su observación especial está aseverando que ella misma está levantando el brazo (por 34 y 36).
- 38) X sabe que la persona que cae bajo su observación especial cree* que ella misma está levantando el brazo.

Así que, en efecto:

- 39) X cree* que él mismo está levantando el brazo.

Una vez más, la persona referida por “X” no puede ignorar la identidad que guarda con la persona referida por “él mismo”. De manera que de 39) se sigue:

- 40) X sabe que la creencia* de que alguien está levantando el brazo es sobre quien cree* que alguien está levantando el brazo, es decir, él mismo.

De modo equivalente, podemos formular 40) como:

- 41) X sabe que él mismo cree* que (él mismo) está levantando el brazo.

²⁵ Parece que, de entrada, Shoemaker no aceptaría que un hablante pueda realizar una aserción sin tener un conocimiento no observacional de que él mismo está realizando la aserción. Pues, como es sabido, siempre ha sido un crítico de cualquier concepción perceptual del autoconocimiento, como el conocimiento de las propias acciones. Ahora bien, es evidente que en este artículo defendiendo un modelo del uso competente de “yo”, el modelo de la autorreferencia autoconsciente, que excluye una concepción perceptual del autoconocimiento. La cuestión es otra: si es concebible que alguien pueda realizar una aserción teniendo sólo un conocimiento observacional o perceptual de lo que está haciendo (Véanse las notas 17 y 21).

Éste es un caso en que, desde luego, mediante la aserción de “Yo estoy levantando el brazo” X expresa la creencia* *consciente* de que él mismo está levantando el brazo: X sabe que la creencia* de que uno mismo está levantando el brazo es *su* creencia y, por lo tanto, X sabe que él mismo cree* eso. ¿Cuál es entonces el factor adicional, a falta del conocimiento de que él mismo es el hablante que realiza la aserción, que nos habilita para llegar a esa conclusión? Desde luego, Shoemaker argumentaría que el factor adicional es que, tanto al realizar la aserción de “Yo estoy levantando el brazo” como al realizar, por ejemplo, la aserción de “Llueve”, la creencia* expresada por la aserción está disponible para X y, en consecuencia, X tiene una creencia sobre la creencia* expresada: en el primer caso, X cree que él mismo cree* que (él mismo) está levantando el brazo. Ahora bien, me parece que, efectivamente, en ambos casos es una condición necesaria de la aserción que la creencia* expresada esté disponible. Pero no veo que de ahí se siga que X tenga una creencia sobre la creencia* expresada, pues es evidente que la creencia* expresada está disponible simplemente en el sentido de que X está dispuesto a comportarse como si lo creyera, específicamente, está dispuesto a realizar la aserción. Esto no entraña en modo alguno que, al realizar la aserción de “Yo estoy levantando el brazo”, esto es, al comportarse como si creyera que él mismo está levantando el brazo, X tenga una creencia sobre la creencia* expresada: la creencia de que se comporta como si creyera que él mismo está levantando el brazo.

El factor adicional que explica que X exprese la creencia* consciente de que él mismo está levantando el brazo es, sin duda, el hecho de que la aserción de “Yo estoy levantando el brazo”, a diferencia de la aserción de “Llueve”, incluye un uso de “yo”. Pues, como hemos visto, del hecho de que X crea* que él mismo está levantando el brazo, que es la creencia* expresada por la aserción mediante un uso de “yo”, se sigue que X sabe que la creencia* de que alguien está levantando el brazo es sobre quien cree* que está levantando el brazo, es decir, él mismo. Así que X sabe que él mismo cree* que (él mismo) está levantando el brazo. De igual modo, como la aserción de la oración mooreana “Llueve pero yo no creo que llueve” incluye un uso de la primera persona, la primera parte de la aserción expresa la creencia* consciente de que llueve: del hecho de que X crea* que él mismo no cree que llueve, que es la creencia* expresada por la segunda parte de la conjunción mediante el uso de “yo”, se sigue que X sabe que la creencia* de que alguien no cree que llueve es sobre quien cree* que llueve, es decir, él mismo. Así que X sabe que él mismo cree* que llueve.

Pero la crítica de Shoemaker que he presentado anteriormente es extensible al caso de una aserción sincera. Es, por supuesto, una condición necesaria de la aserción sincera de “Yo estoy levantando el brazo”, que la creencia expresada esté disponible para X. Pero puede sostenerse, como el propio Shoemaker lo hace, que la creencia expresada está disponible simplemente en el sentido de que si X, entre otras cosas, se pregunta si el contenido de la creencia es o no es verdadero, entonces realizará la aserción sincera de ese contenido.²⁶ X realiza la aserción sincera de “Yo estoy levantando el brazo” porque, al preguntarse si es o no es verdad que él mismo está levantando el brazo, la creencia de que él mismo está levantando el brazo está disponible y, por ello, puede responder que efectivamente él mismo está levantando el brazo. Sin embargo, X no está preguntándose directamente si tiene o no tiene la *creencia* de que él mismo está levantando el brazo sino que está preguntándose si es o no es verdad que él mismo está levantando el brazo. Ciertamente, si X está preguntándose si es o no es verdad que él mismo está levantando el brazo y posee el concepto de creencia, entonces está en condiciones de preguntarse si tiene alguna creencia al respecto. Pero no tiene que estar haciendo eso, averiguando si tiene o no tiene la creencia. No es necesario por tanto que, al estar disponible la creencia expresada, X crea que él mismo cree que (él mismo) está levantando el brazo. La conclusión es que también la aserción sincera de la oración mooreana “Llueve pero yo no creo que llueve” expresa, en la primera parte de la conjunción, la creencia consciente de que llueve porque la aserción completa incluye un uso de “yo”: del hecho de que X crea que él mismo no cree que llueve, que es la creencia expresada por la segunda parte de la conjunción mediante el uso de “yo”, se sigue que X sabe que la creencia de que alguien no cree que llueve es sobre quien cree que llueve, es decir, él mismo. Así que X sabe que él mismo cree que llueve.

De manera que la explicación del sinsentido o irracionalidad propio de una aserción mooreana no precisa recurrir, como Shoemaker lo hace, a la existencia de una relación constitutiva entre creencia (o creencia*) y conciencia. Concluyo, entonces, que la explicación de la paradoja de Moore basada en la naturaleza de la primera persona tiene prioridad sobre la explicación basada en la naturaleza consciente de la creencia. Más aún, me parece que el análisis anterior introduce una razón para pensar que la relación constitutiva no es entre creen-

²⁶ O, al menos, realizará un acto de asentimiento mental. En lo que sigue daré por supuesta esta aclaración.

cia (o creencia*) y conciencia, sino entre creencia (o creencia*) de *primera persona* y conciencia. Pues, olvidándonos por un momento de la aserción, tenemos ahora una razón para pensar que a la creencia misma atribuida en:

42) X cree que llueve

le falta el factor relevante que pudiera justificar la existencia de una relación *constitutiva* con la creencia de segundo orden atribuida en:

43) X cree que él mismo cree que llueve.

Como he argumentado, aunque la creencia descrita en 42) esté disponible, de ahí no se sigue que X tenga la creencia descrita en 43). En efecto, a la creencia disponible descrita en 42) le faltaría el factor de la primera persona que justifica la existencia de una relación constitutiva entre la creencia de primera persona atribuida en:

44) X cree que él mismo está levantando el brazo

y la creencia de segundo orden atribuida en:

45) X cree que él mismo cree que (él mismo) está levantando el brazo.

Desde luego, una consecuencia de tener la creencia descrita en 44) es que esa creencia está disponible para X, en el sentido de que si X se pregunta si el contenido de la creencia es o no es verdadero, realizará la aserción de “Yo estoy levantando el brazo”. No puede ocurrir que esa creencia no esté disponible. Pero no es que X tenga la creencia de segundo orden descrita en 45) porque la creencia descrita en 44) esté disponible: 45) se sigue de 44) debido al hecho de que la creencia descrita en 44) es una creencia de primera persona y, por lo tanto, X tiene la creencia de que él mismo cree que (él mismo) está levantando el brazo.²⁷

Ésta es una razón para concluir también que la explicación de la paradoja de Moore en términos de la naturaleza consciente de la creencia finalmente debe ser desestimada. Desde luego, el hablante que realiza la aserción sincera de la oración “Llueve pero yo no

²⁷ Como he señalado en varias ocasiones, prefiero hablar de tener un conocimiento, que no es necesariamente una creencia, que de tener una creencia de segundo orden:

45*) X sabe que él mismo cree que (él mismo) está levantando el brazo.

creo que llueve” tendrá una creencia de segundo orden, o algún tipo de conocimiento, sobre la creencia de que llueve, que es la creencia expresada por la primera parte de la conjunción y que, como Shoemaker sostiene, está disponible. Entonces, la creencia de segundo orden también estará disponible y él estará dispuesto a realizar la aserción sincera de la oración “Yo creo que llueve”, cuando de hecho está aseverando sinceramente su negación mediante la segunda parte de la conjunción mooreana. Pero la explicación no está en que, al realizar la aserción mooreana, la creencia de que llueve esté disponible y, por consiguiente, sea una exigencia racional que el hablante tenga una creencia de segundo orden sobre ella. La explicación está en que, como la aserción mooreana incluye un uso del pronombre personal “yo”, la creencia compuesta expresada por la aserción es de primera persona y, por ello, el hablante tiene una creencia de segundo orden, o algún tipo de conocimiento, sobre la creencia de que llueve.

VI

He argumentado que una explicación de la paradoja de Moore en términos de la naturaleza de la primera persona es independiente de otras explicaciones basadas en la naturaleza consciente de la aserción o en la naturaleza consciente de la creencia.²⁸ Pues, por un lado, he mostrado que hay casos igualmente paradójicos, cuyo sinsentido o irracionalidad es explicable en términos del uso del pronombre personal “yo”, pero en los que, sin embargo, la naturaleza consciente de la aserción mooreana es explicativamente insuficiente. Por otro lado, establezco que para explicar la paradoja de Moore no es necesario suponer, teniendo una explicación en términos del uso de “yo” en la aserción mooreana, que exista una relación constitutiva

²⁸ Añado finalmente que he examinado la paradoja de Moore como el tipo de absurdo que resulta de *aseverar* el contenido proposicional de una oración mooreana. En la literatura al respecto suele distinguirse otra paradoja de Moore: el tipo de absurdo que resulta de *creer* el contenido proposicional de una oración mooreana (Fernández 2005, pp. 534–536). No puedo ocuparme aquí del alcance de la explicación en términos de la primera persona en relación con este segundo fenómeno. Sólo diré que si es constitutivo de tener una creencia, como un tipo de suceso que está ocurriendo en la mente, que haya también un acto o acción mental de juzgar, entonces el modelo de la autorreferencia autoconsciente puede aplicarse igualmente a la autoconciencia de un creyente: el creyente sabe que él mismo realiza la acción mental y también conoce la regla de autorreferencia para el concepto de primera persona que forma parte de la creencia, en cuyo caso sabe que su concepto de primera persona refiere a él mismo (Peacocke 2008, pp. 78–83; O’Brien 2007, pp. 89–94). En ese caso, es esperable que una explicación en términos de la primera persona pueda darnos una explicación *uniforme* de las paradojas lingüística y psicológica.

entre creencia y conciencia. Esto significa, en primer lugar, que la explicación que he presentado en términos de la primera persona tiene cierta prioridad sobre las otras explicaciones. Pero es que, además, he sostenido que el análisis anterior nos entrega razones para pensar que la aserción no es de naturaleza consciente y que no existe una relación constitutiva entre creencia y conciencia. Como resultado de ello deberíamos concluir que las otras explicaciones ya no son candidatos en juego.²⁹

BIBLIOGRAFÍA

- Anscombe, G.E.M., 1981, "The First Person", *Metaphysics and the Philosophy of Mind*, Basil Blackwell, Oxford, pp. 21–36.
- Bermúdez, J.L., 1998, *The Paradox of Self-Consciousness*, The MIT Press, Cambridge Mass.
- Castañeda, H.N., 1999, "On the Phenomeno-Logic of the I", *The Phenomeno-Logic of the I*, Indiana University Press, Bloomington, pp. 89–95.
- Fernández, J., 2005, "Self-Knowledge, Rationality and Moore's Paradox", *Philosophy and Phenomenological Research*, vol. 21, no. 3, pp. 533–556.
- Moore, G.E., 1944, "Russell's Theory of Descriptions", en P. Schlipp (comp.), *The Philosophy of G.E. Moore*, Tudor, Evanston, pp. 175–225.
- O'Brien, L., 2007, *Self-Knowing Agents*, Oxford University Press, Oxford.
- Peacocke, C., 2008, *Truly Understood*, Oxford University Press, Oxford.
- , 1983, *Sense and Content*, Oxford University Press, Oxford.
- Rosenthal, D.M., 2010, "Expressing One's Mind", *Acta Analytica*, vol. 25, pp. 21–34.
- , 2002, "Moore's Paradox and Crimmins' Case", *Analysis*, vol. 62, pp. 167–171.
- , 1995, "Moore's Paradox and Consciousness", *Philosophical Perspectives*, vol. 9, pp. 313–333.
- Shoemaker, S., 2009, "Self-Intimation and Second Order Belief", *Erkenntnis*, vol. 71, pp. 35–51.
- , 1996, "Moore's Paradox and Self-Knowledge", *The First-Person Perspective and Other Essays*, Cambridge University Press, Cambridge, pp. 74–93.
- , 1994, "Self-Reference and Self-Awareness", en Q. Cassam (comp.), *Self-Knowledge*, Oxford University Press, Oxford, pp. 80–93.

²⁹ Agradezco al profesor Julio Torres, del Departamento de Filosofía de la Universidad de Concepción, y a los árbitros de esta revista los comentarios y sugerencias que me han hecho, que, sin duda, contribuyeron a mejorar la calidad del artículo.

- Vidal, J., 2011, “La autorreferencia autoconsciente y el uso de ‘yo’ como sujeto”, *Analítica*, vol. 5, pp. 61–79.
- Williams, J.N., 2006, “Wittgenstein, Moorean Absurdity and Its Disappearance from Speech”, *Synthese*, vol. 149, pp. 225–254.
- Williamson, T., 2000, *Knowledge and Its Limits*, Oxford University Press, Oxford.

Recibido el 19 de julio de 2011; revisado el 18 de marzo de 2012; aceptado el 26 de marzo de 2012.