LA VOLUNTAD DE NO CREER

MANUEL HERNÁNDEZ IGLESIAS
Departamento de Filosofía. Universidad de Murcia.

Resumen

¿Puede ser racional creer algo porque se quiere creerlo? *Prima facie* no, puesto que una creencia racional se basa en razones, y la voluntad de creer puede ser una causa, pero no una razón de la creencia. Sin embargo, algunos intentos de autoinducirse creencias causalmente pueden verse como el ejercicio de una racionalidad de segundo orden. En este artículo, se esboza una visión de esta racionalidad de segundo orden basada en la tesis de Davidson de la división de la mente y la distinción de Ortega entre ideas y creencias. La inquietante conclusión del análisis es que la capacidad de tener deseos de segundo orden sobre nuestras creencias es lo que hace posibles tanto la autocrítica y la autosuperación como el dogmatismo y la autocorrupción y que la condición de posibilidad de la racionalidad y la libertad es la misma que la de la irracionalidad y la alienación.

PALABRAS CLAVE: Racionalidad - voluntad - creencias - autoengaño - mente dividida.

Abstract

Is it rational to believe something because one wants to believe it? *Prima facie* it is not, for a rational belief is based on reasons, and the will to believe can be a cause, but not a reason for the belief. Nevertheless, some attempts of causally self-inducing beliefs can be viewed as a sort of second-order rationality. This paper sketches a view of this second-order rationality based on Davidson's thesis of the divided mind and Ortega's distinction between ideas and beliefs. The uncomfortable conclusion of the discussion is that the ability to have second-order beliefs about our beliefs is what makes possible both self-criticism and self-improvement and dogmatism and self-corruption and that the possibility condition of rationality and freedom and of irrationality and alienation are the same.

KEY WORDS: Rationality - will - beliefs - self-deception - divided mind.

La fe salva, *luego* es falsa.
Nietzsche, *El Anticristo*

1. Creencias y deseos

El punto de partida de este trabajo es la cuestión de si un sujeto puede creer algo voluntariamente. Esta pregunta es sin embargo ambigua, y dependiendo de cómo se entienda, la respuesta será muy diferente.

Si el "puede" se interpreta en sentido normativo, es decir, como la pregunta de si es epistemológicamente legítimo creer algo voluntariamente o si puede estar justificado creer algo voluntariamente, la respuesta

parece claramente negativa. Las propias expresiones de la psicología popular que describen estos fenómenos de implicación de la voluntad de creer en las creencias mismas tienen una carga valorativa negativa: "*confundir* los deseos con la realidad", "*engañar*se a uno mismo", etc. Tanto la psicología popular como la lógica informal o la teoría de la argumentación consideran las intervenciones de la voluntad de creer en la adopción de las creencias como interferencias que dificultan el buen juicio.

La manifestación más clara de que la voluntad de creer algo no puede ser una buena razón para creerlo es el hecho de que la afirmación de que uno mismo cree algo porque quiere creerlo es paradójica. Por ejemplo,

(1) Yo creo que *p* porque quiero creerlo (o: *p*, porque quiero creerlo)

es una contradicción pragmática del mismo tipo que la paradoja de Moore ("*p*, pero no creo que *p*"). En ambos casos, la paradoja deriva de suponer que el hablante está respetando las máximas conversacionales, concretamente las máximas de calidad. En el caso de la paradoja de Moore, si suponemos que el hablante respeta el Principio Cooperativo y, por consiguiente, la primera máxima de calidad, hemos de suponer que no afirma aquello que considera falso. De ahí la contradicción pragmática entre afirmar en un mismo enunciado que *p* y que uno no cree que *p*.

¿Por qué (1) nos parece una contradicción pragmática? Por la misma razón que la paradoja de Moore, sólo que en este caso está implicada la segunda máxima de calidad, no la primera. Dicha máxima establece que no debe afirmarse aquello de lo que se carezca de pruebas adecuadas. Por ello, afirmar que se cree que *p* al tiempo que se niega que dicha creencia se base en prueba alguna resulta pragmáticamente contradictorio. Salvo, claro está, que se considere que la voluntad de creer una proposición puede contar como prueba de ésta. Es decir, salvo que se niegue que las cosas son como son, y no necesariamente como uno quisiera que fueran.[1]

[1] De lo anterior no se sigue ni mucho menos que todo tipo de intervención de la voluntad en la formación de las creencias sea una patología epistémica. Es obvio que los deseos e intereses intervienen necesaria y legítimamente en la adquisición de conocimiento. Y no sólo determinando el objeto del estudio, la reflexión o la investigación. También determinando las tesis que se va a intentar argumentar o probar. Lo único que se excluye por incompatible con su propia percepción es la intervención de la voluntad, los deseos o los intereses como parte de la justificación de las creencias.

Una posible objeción a lo anterior es que, puesto que (1) es una contradicción pragmática, lo que sucedería no es que el sujeto cree que *p* por motivos irracionales, sino que en realidad no cree que *p*. Esto es cierto, pero no demuestra que el sujeto no pueda tener una creencia porque desea creerla, sino que es imposible que tenga una cre-

Si, por el contrario, entendemos el "puede" de la pregunta en un sentido descriptivo, la respuesta es que, obviamente, sí. Los sujetos no sólo pueden, sino con frecuencia creen cosas al menos en parte debido a su deseo de creerlas, o dejan de creer cosas por su deseo de no creerlas. Fenómenos como el pensamiento desiderativo o el autoengaño pueden ser difíciles de definir y analizar, pero es un hecho que todos apelamos a ellos a menudo para explicar por qué alguien tiene tal creencia o tal conjunto de creencias. Es más, en muchas ocasiones, la atribución de confusión entre deseos y realidad o de autoengaño son las únicas explicaciones que nos permiten dar sentido a algunos comportamientos. Forma parte esencial de nuestra comprensión del otro la apelación a estos mecanismos, puesto que, en muchos casos, prescindir de ellos haría que muchas creencias o conductas parecieran irracionales en el sentido radical de resultar incomprensibles.

El pensamiento desiderativo o el autoengaño no los atribuimos sólo a los demás, sino también a nuestro propio comportamiento pasado. Y no sólo pasado, porque la posibilidad de que las propias creencias estén al menos parcialmente motivadas por alguna forma de autoengaño o pensamiento desiderativo es una hipótesis que una persona mínimamente juiciosa no puede dejar de tener en cuenta. Obviamente, nadie es más proclive a caer en formas desiderativas de pensamiento que quien se considera inmune a ellas, como sucede con cualquier otra posible perversión del razonamiento.

Por tanto, interpretada descriptivamente, la pregunta inicial parece tener una clara respuesta: es un hecho psicológico (tanto de la psicología científica como de la popular) que, frecuentemente, la voluntad del sujeto de creer algo es parte esencial de la explicación del hecho de que efectivamente lo crea, ya sea directamente (caso del pensamiento desiderativo), ya sea indirectamente por medio de alguna forma de autoengaño. De ahí que, a diferencia de (1), las oraciones siguientes no sean pragmáticamente paradójicas:

(2) X cree que *p* porque quiere creerlo.
(3) Yo creía que *p* porque quería creerlo.

---

encia porque desea tenerla y ser consciente de ello. Puede considerarse que el tener razones adecuadas para creer que *p* es un requisito para creer que *p* y que, por lo tanto, lo que sucede es que el sujeto cree tener una creencia que no tiene. Posiblemente hay situaciones cuya descripción más adecuada sea ésa, pero generalizarlo a todos los casos llevaría a la conclusión de que las creencias no basadas en razones legítimas no son creencias que uno tiene, sino, a lo sumo, creencias que uno cree tener. Y esto me parece una restricción demasiado arbitraria del uso del término "creencia".

En conclusión, la voluntad de creer (o no creer) que *p*, sea lo que sea, no puede ser una justificación de *p*, no puede formar parte de las razones. Sí puede ser parte de la explicación de por qué un sujeto cree que *p*, pero no de las razones a favor de *p*. No es que sea una mala razón, es que no es una razón en absoluto, porque presentarla, en primera persona del presente del indicativo, como una razón de la propia creencia es pragmáticamente paradójico. No es sin embargo paradójico en segunda o tercera personas, ni en otros tiempos verbales, porque sí es un hecho que el deseo de creer algo puede dar y a menudo da lugar a que alguien crea ese algo.

2. Las paradojas de la irracionalidad

Como se ha dicho más arriba, los enunciados (2) y (3), a diferencia de (1), no son paradojas pragmáticas en el sentido ejemplificado por la paradoja de Moore. La razón es que lo que resulta inmediatamente paradójico es sólo el autoengaño o pensamiento desiderativo conscientes. Mientras el sujeto ignore que su deseo de creer que *p* es una de las razones esenciales por las que cree que *p*, el sujeto no se contradice (o al menos no de manera flagrante).

No obstante, las afirmaciones de que un sujeto cree algo porque desea creerlo, aun formuladas en tercera persona o en un tiempo verbal distinto del presente del indicativo, aunque no inmediatamente paradójicas, sí plantean lo que Davidson ha denominado paradojas de la irracionalidad.[2]

Explicado brevemente, lo paradójico de enunciados como (2) y (3) es que afirman algo al tiempo que implican lo contrario. Concretamente, afirman explícitamente que alguien cree que *p*, al tiempo que implican que no cree que *p* de la manera siguiente: si el deseo de S de creer que *p* es parte esencial de la explicación de la creencia, entonces el resto de las creencias de S son insuficientes para justificar la creencia de que *p*. Pero, puesto que el deseo de creer que *p* no es una razón para creer que *p*, hay que admitir también que, en cierto sentido, S no cree propiamente que *p*. No al menos si, con Davidson, admitimos el holismo y la necesaria atribución a los sujetos de pautas racionales de razonamiento como condición de posibilidad de la interpretación. Si la atribución de estados intencionales está, como Davidson insiste, sometida a normas de racionalidad, parece que es incompatible atribuir a un sujeto estados a la vez irracionales e intencionales.

[2] Los textos de Davidson sobre la irracionalidad son Davidson, 1982, 1985, 1986 y 1997.

Davidson resuelve esta paradoja distinguiendo, en los estados psicológicos que son causas de otros, entre aquellos que además de causas son razones y aquellos que sólo son causas.[3] En nuestro caso, el estado psicológico consistente en desear creer que *p* actúa como causa (o parte esencial de la causa, o una de las causas) de la creencia de que *p*, aunque no es una razón para creer que *p*. Como Davidson señala, la naturaleza holista del pensamiento implica que, cuando un estado psicológico actúa como causa de otro estado psicológico para el cual no constituye una razón, ha de tener lugar una cierta escisión del sujeto, lo que él llama una mente dividida. Sólo en una mente dividida puede suceder que un estado psicológico cause otro para el que no es ni puede ser una razón.[4]

Esto explica por qué (1) es una contradicción pragmática: el sujeto no puede conscientemente creer que *p* al menos en parte por su deseo de creerlo porque, si es consciente de ello, su mente, en lo que a la creencia de que *p* se refiere, no está dividida. Para que el sujeto pueda creer que *p* por su deseo de creerlo, es necesario que haya una separación entre la parte de su mente que cree que *p* y la parte de su mente que desea creer que *p*. Algo que claramente no se da en el sujeto que emite (1). Y también explica por qué (2) y (3), a diferencia de (1), no nos parecen paradójicas, pues el "porque" que en ellas aparece es claramente causal, no racional.

Esta idea de la división de la mente permite a Davidson atribuir a los estados psicológicos de un mismo individuo el tipo de relaciones causales no racionalizadoras que se dan entre estados psicológicos de dos individuos distintos, sólo que, en este caso, en lugar de mentes distintas, lo que tendríamos serían distintos compartimentos de una misma mente.[5] Este planteamiento armoniza muy bien con el lenguaje de la psicología popular, haciendo inteligibles expresiones como la de "engañarse a uno mismo", que usamos y comprendemos sin dificultad, pero cuyo sentido parece imposible de explicar (si uno miente, uno sabe que miente, y por lo tanto, no puede uno engañarse). La hipótesis de una mente dividida ofrece al menos un esquema de explicación de este aparentemente paradójico fenómeno: es un compartimento de la mente el que causa la creencia en otro compartimento, y la compartimentación hace que esta relación causal permanezca invisible para el segundo de ellos.

Por supuesto, la noción de un yo dividido requiere más elaboración, algo que ni me propongo ni me considero capacitado para hacer. Así que, en adelante, voy simplemente a presuponer que la noción está suficien-

[3] Habría que añadir los que, siendo razones, no actúan como causas.
[4] O que un estado psicológico del sujeto que es una razón no actúe como causa.
[5] Véase el ejemplo de Davidson del jardinero (Davidson, 1982, p. 181).

temente clara para mis propósitos o que, al menos, puede clarificarse suficientemente. La única precisión que deseo hacer es que la división de la mente exigida por el análisis de Davidson de la irracionalidad es una división más fuerte que la mera no omnisciencia lógica.

3. La autotrascendencia

Podemos resumir lo dicho hasta ahora del modo siguiente. Se ha argumentado, en la primera sección, que el deseo de creer que *p* no puede ser una razón para creer que *p*, por lo que un sujeto no puede creer algo porque desea creerlo y ser consciente de ello. Sí puede, como se ha expuesto en la segunda sección, creer que *p* porque desea creer que *p*, pero sólo en el caso de una mente dividida en la que el deseo actúa como causa y no como razón de la creencia.

De lo anterior, parece seguirse que creer algo, al menos en parte, por el deseo de creerlo, es irracional, puesto que la creencia dependería esencialmente de estados psicológicos que no son razones para la creencia. Sin embargo, esta conclusión puede resultar o no precipitada, dependiendo de lo evidente que se considere que es irracional toda creencia causada por estados psicológicos que no son razones para ella. Esta tesis, la de la irracionalidad de una creencia generada (al menos en parte) por causas que no son razones, es la que discutiré en el resto del artículo.

Si admitiéramos la tesis en cuestión, estaríamos en situación de dar por zanjada la pregunta de si puede, en el sentido normativo de "puede", creerse algo porque se desea creerlo. La respuesta sería claramente negativa, puesto que la relación entre el deseo y la creencia es puramente causal, no racional. Lo cual confirmaría nuestra intuición inicial.

Pero, ¿es esto así?, ¿es necesariamente irracional provocarse a uno mismo creencias causalmente, en lugar de fundamentarlas en razones? *Prima facie* parece claro que sí. Si una creencia racional es una creencia basada en razones (¿y qué otra cosa podría ser?), sólo podemos reconocer como racional una creencia causada por estados psicológicos que actúen como razones para la creencia, aunque sean malas razones (la racionalidad no implica la omnisciencia ni material ni lógica).[6] Pero no podemos

[6] Distinto es el caso de la apelación a razones que parecen malas al propio sujeto. En este supuesto, diríamos que el sujeto no actúa realmente por dichas razones. A lo sumo, afirma actuar por ellas, en cuyo caso la justificación de su acción puede calificarse de hipócrita (si lo que hace es lo que en inglés se denomina "*to give the right reason*"). Pero la hipocresía de su justificación (de las razones que ofrece para su acción) no convierte *per se* la acción misma en irracional. Es más, del mismo modo que, como afirmaba Voltaire, la hipocresía

reconocer como racional la adopción de una creencia como resultado de algo que no es una razón en absoluto, ni buena ni mala. Esto se manifiesta en que, por ejemplo,

(4) Creo que *p* porque C, pero C no es una razón para *p*

es una contradicción pragmática por las mismas razones que (1).

No obstante, hay un sugerente párrafo de Davidson que apunta en el sentido opuesto al razonamiento anterior. En él, Davidson defiende la racionalidad de modificar los propios estados psicológicos, no por medio de razones, sino actuando causalmente sobre ellos. Y no sólo eso, sino que considera tal tipo de modificación, no sólo racional, sino particularmente valioso. El párrafo, que cito *in extenso*, es el siguiente:

> ¿Pero proporciona el esquema [de explicación de la irracionalidad] una condición suficiente de la irracionalidad? Parecería que no. Pues los casos sencillos de asociación no cuentan como irracionales. Si consigo recordar un nombre tarareando cierta melodía, hay una causa mental de algo para lo cual no es una razón; y lo mismo para multitud de otros casos. Pero mucho más interesante, y más importante, es un tipo de autocrítica y perfeccionamiento que tendemos a tener en alta estima, y que incluso ha sido considerado la esencia misma de la racionalidad y la fuente de la libertad. No obstante, es un caso claro de causalidad mental que trasciende la razón (en el sentido de alguna manera técnico en que he venido usando el concepto).
>
> Lo que tengo en mente son un tipo especial de deseo o valor de segundo orden y las acciones que éste puede provocar. Esto sucede cuando una persona se forma un juicio positivo o negativo sobre algunos de sus propios deseos, y actúa para cambiar dichos deseos. Desde el punto de vista del deseo cambiado, no hay razón para el cambio (la razón viene de una fuente independiente y se basa en consideraciones adicionales y en parte contrarias). El agente tiene razones para cambiar sus propios hábitos y carácter, pero estas razones vienen de un dominio de valores necesariamente extrínseco a los contenidos de las opiniones o valores que sufren el cambio. La causa del cambio, si se produce, puede por lo tanto no ser una razón para lo que causa. Una teoría que no pudiera explicar la irracionalidad sería tal que tampo-

es el homenaje que el vicio rinde a la virtud, la argumentación hipócrita es un homenaje que la irracionalidad rinde a la razón, en el sentido de que el que argumenta hipócritamente al menos reconoce el espacio de las razones como el legítimo para la toma de decisiones.

> co podría explicar nuestros saludables esfuerzos, y ocasionales éxitos, de autocrítica y autosuperación. (Davidson, 1982, pp. 186-7)

Aplicadas al ámbito de las creencias, no al de los deseos, estas consideraciones sugieren la distinción entre dos tipos de racionalidad epistémica. Una sería la racionalidad de primer orden, que consiste en adoptar creencias cuyas causas sean todas ellas razones, evitando la intervención exclusivamente causal de deseos o intereses en la adopción de la creencia. La otra sería la racionalidad de segundo orden, que consiste en la autoinducción de creencias por medio de la provocación **deliberada** de situaciones susceptibles de actuar como causas, aunque no como razones, de dichas creencias. Llamaremos a la racionalidad de primer orden objetividad y a la de segundo orden, siguiendo a Marcia Cavell (Cavell, 1999, p. 408), autotrascendencia. La diferencia entre una y otra es que la autotrascendencia, a diferencia de la objetividad, implica la autopercepción del sujeto como una mente dividida.

El problema es encontrar una caracterización adecuada del papel de la voluntad en la autotrascendencia epistémica. Dicha caracterización deberá cumplir el requisito de no ser reductible al mero esfuerzo por eliminar la influencia de deseos o intereses en la formación de las creencias. En consecuencia, **debe involucrar realmente el intento de provocar la adopción de estados psicológicos que, no siendo razones para la creencia que se desea tener, sí tiendan a causarla. Ello implica que el sujeto debe asumir una cierta escisión de su propia mente y, en cierto sentido, explotarla, sacarle partido**.

De nuevo, siguiendo a Davidson, podemos recurrir a la analogía entre las relaciones causales entre compartimentos de una misma mente y las relaciones causales entre estados psicológicos de mentes distintas. El caso de la autotrascendencia tendría una similitud con el caso de una persona racional que quisiera inducir una creencia en otra de cuya racionalidad tiene una opinión muy pobre. Consciente de la escasa disposición de su interlocutor para atender a las razones que justifican la creencia, la primera procuraría provocar en la segunda los estados psicológicos que tiendan a causar la creencia que él desea que adopte, aunque dichos estados no sean razones para ella.

En el caso de la autotrascendencia sería uno de los compartimentos de la mente el que trataría a otro de ellos como una especie de menor de edad racional. La autotrascendencia, así entendida, implicaría:

1. La autopercepción del sujeto como una mente dividida en la que ciertos estados psicológicos tienden a causar creencias para las que no son razones.

2. La voluntad de autoprovocarse estados psicológicos que tienden a causar creencias deseadas para las que no son razones (o la voluntad de eliminar los estados psicológicos que tienden a causar creencias no deseadas para las que no son razones).

La condición de que los estados psicológicos que el sujeto desea provocarse a sí mismo no sean razones para las creencias es necesaria para que la caracterización de la autotrascendencia satisfaga el requisito establecido más arriba de no reductibilidad a la mera voluntad de objetividad (entendida como la mera eliminación de causas que no sean razones).

En el párrafo de Davidson citado más arriba, el proceso parte de una situación en la que el sujeto se forma una opinión negativa de alguno de sus deseos. En el caso epistémico, que es el que nos ocupa, la situación equivalente sería la de un sujeto que se forma una opinión negativa con respecto a alguna de sus creencias. **La cuestión es, ¿qué tipo de consideraciones justifican que uno se forme una opinión negativa de una de sus creencias?** La candidata más obvia para justificar la opinión negativa acerca de una creencia *p* es, por supuesto, la creencia de que *p* es falsa, pero en este caso lo que sucede es que el sujeto no tiene la creencia de que *p*, por lo que no necesitaría extirpársela. Otras candidatas obvias son otras creencias que sean buenas razones en contra de *p*, pero en este caso no habría autotrascendencia, puesto que las creencias actuarían como razones de la creencia de que no *p*, no como meras causas de ella. Así que lo que buscamos son consideraciones que, no siendo razones en contra de la creencia, justifiquen no obstante una opinión negativa sobre ella y proporcionen un motivo racional para desear abandonarla.

4. Autotrascendencia y autocorrupción

El mejor modo de hacer visible esta dificultad es comparando la autotrascendencia con su contrapartida negativa. En un texto posterior, Davidson se refiere al fenómeno que hemos denominado autotrascendencia con una mayor desconfianza:

> Hacer o pensar cosas con el objetivo consciente o inconsciente de cambiar nuestras propias creencias o actitudes proposicionales no es necesariamente malo, ni siquiera lo que normalmente llamaríamos irracional. John Dewey, que, en la línea de Aristóteles, era pesimista acerca de la posibilidad de hacer gran cosa para cambiar los propios valores, habló hace muchos años sobre cómo, con suerte y esfuerzo, puede hacerse (*Human Nature and Conduct*). Su propuesta tenía dos par-

> tes: la primera era que si quieres tener un valor o una creencia que no tienes, deberías actuar como si ya los tuvieras. La segunda parte era evitar concentrarse en el fin deseado y concentrarse en los medios. No sigas repitiéndote "No fumaré", sino prepara una excursión interesante hacia donde no se pueden encontrar cigarrillos. Dewey no advirtió que su consejo funciona mejor al servicio de la autocorrupción. Si tu deseo secreto es cometer adulterio, no te digas "Me dejaré seducir"; simplemente déjale tocarte la mano. (Davidson, 1999, p. 229)

El caso de la autocorrupción epistémica es paralelo al de la autotrascendencia. En ambos, el sujeto es consciente de tener una mente dividida en la que ciertos estados psicológicos causan creencias para las que no son razones. Y, en consecuencia, adopta un comportamiento orientado a provocar en él el tipo de estados psicológicos que tienden a causar la creencia deseada y evitar los que tienden a causar la creencia no deseada. Es decir, en el caso de la autocorrupción epistémica, se cumplen las dos condiciones con las que caracterizamos su variante positiva, la autotrascendencia.

En los dos casos, el sujeto parte del deseo de adoptar una creencia para la que no encuentra razones. Esta impotencia le lleva a adoptar hacia sí mismo la perspectiva de una tercera persona que intenta inducirle causalmente las creencias deseadas. ¿Cuál es la diferencia entre ambos? Si, como hemos visto, el proceso es formalmente idéntico, la diferencia radicará en el tipo de motivación por la que un sujeto desea tener una creencia o dejar de tenerla; en el origen de su incomodidad ante una creencia que tiene y no quiere tener o ante la ausencia de una creencia que desearía tener.

Son muchos los motivos por los que uno puede desear tener o dejar de tener una creencia, y no me propongo hacer un inventario. Pero, siguiendo a Bernard Williams (Williams, 1970, p. 149 y ss.), es importante dividir estos motivos en dos categorías: los centrados en la verdad y los no centrados en la verdad. Un motivo centrado en la verdad para desear creer que p es aquél que sólo encuentra satisfacción si la creencia de que p es una creencia verdadera. Es decir, si el deseo de creer que p está vinculado al deseo de que p. Un motivo no centrado en la verdad para desear creer que p sería un deseo desvinculado del deseo de que p. Es decir, sería el deseo de creer que p, con independencia de si p o no p.

Williams propone el ejemplo de un padre que desea creer que su hijo, que ha desaparecido en el mar tras un accidente, está vivo. Este deseo está centrado en la verdad porque no se vería satisfecho si la creencia se produjera por un procedimiento absolutamente desligado de su verdad;

por ejemplo, el de acudir a un hipnotizador que le causara la creencia. El padre, a pesar de su deseo de creer, no acudiría al hipnotizador porque eso no daría satisfacción a su deseo de creer que el hijo está vivo. Y eso es así porque quiere creer que su hijo está vivo y que la causa de esa creencia sea que su hijo, en efecto, esté vivo. El caso del deseo no centrado en la verdad sería el del padre que, en esta situación, sí optara por acudir al hipnotizador. En este segundo supuesto, el deseo de creer que el hijo está vivo es un deseo independiente del valor de verdad de la creencia (éste sería el caso si, por ejemplo, lo que deseara el padre es sólo liberarse de la angustia que le produce la incertidumbre sobre la vida de su hijo o la creencia de que su hijo está muerto).

Williams considera que la autoinducción de la creencia en el segundo supuesto es profundamente irracional. Pero, aunque irracional, es coherente con el deseo. Y, si esto es así, entonces es que el propio deseo es irracional. El deseo de creer algo sería racional sólo en el supuesto de que el procedimiento por el que se desea alcanzar la creencia fuera un procedimiento que garantizara la verdad de ésta. En apoyo de esta tesis, Williams apela, en primer lugar, a nuestro rechazo intuitivo al procedimiento de inducción de la creencia del segundo ejemplo. Y, en segundo lugar, y como explicación de este rechazo, apela al carácter holista de nuestra mente. Este carácter holista hace que uno no pueda desprenderse de una creencia sin modificar muchas otras, que a su vez obligarían a modificar otras muchas, con lo que "pudiera ser que un proyecto de este tipo tendiera finalmente a implicar la destrucción total del mundo real, a llevar a la paranoia" (*Ibíd*., p. 151).

De lo anterior, se sigue que lo que hemos llamado autocorrupción epistémica es irracional, puesto que en ella el deseo de inducirse una creencia es autónomo con respecto al deseo de que la creencia sea verdadera. La voluntad actúa en el sentido de eliminar deliberadamente la objetividad. El sujeto decide, por así decirlo, nublar sus propios juicios. En la autocorrupción, el sujeto se autoaplica una especie de ingeniería epistémica en la que abdica de la propia racionalidad al sabotear deliberadamente la objetividad. Con ello, el sujeto, consciente de la división de su mente, abusa de ella y la explota para afianzarse en creencias que sabe que no tiene buenas razones para sostener.

El problema aquí es que, en el caso de la autotrascendencia epistémica, la situación no es muy diferente. Recordemos que tanto la autotrascendencia como la autocorrupción epistémicas partían del deseo de adoptar o rechazar creencias. En ambos supuestos, el sujeto se autoexponía voluntariamente a ciertas situaciones con la intención de que éstas dieran lugar a estados psicológicos que a su vez causaran la adopción o

el abandono de ciertas creencias. Y, en ambos casos, los estados psicológicos que se espera causen las creencias no son razones para ellas.

La autotrascendencia, tal y como la hemos caracterizado, no es un deseo de creer centrado en la verdad como el primero de los de Williams. En éste, el deseo del padre del náufrago de creer que su hijo está vivo sólo puede hallar satisfacción si las causas que producen tal creencia son también razones para ella. Pero, por definición, en la autotrascendencia el sujeto se autoinduce la creencia a través de causas que no son razones. El caso del padre del náufrago no es pues un ejemplo de autotrascendencia. ¿Hemos de concluir que, en el caso de la autotrascendencia, el deseo de modificar las propias creencias es un deseo no centrado en la verdad? Si así fuera, no parece que podamos atribuirle una racionalidad mayor que al padre que acude al hipnotizador para que le provoque artificialmente la creencia de que su hijo está vivo. Mucho menos que pueda ser elevado por nadie a la categoría, nada menos, de "esencia misma de la racionalidad" y "fuente de la libertad".

Lo que hemos llamado autotrascendencia, como algo no sólo distinto, sino opuesto epistémica y éticamente a la autocorrupción, tiene que involucrar de algún modo, la verdad de la creencia deseada. En caso contrario, es decir, si la verdad de la creencia no está implicada de alguna manera en la motivación del deseo de creerla, no puede establecerse la diferencia entre autotrascendencia y autocorrupción. A lo sumo, la autotrascendencia sería un término pomposo para formas benignas de autoengaño. Sería, en el mejor de los casos, merecedora de cierta comprensión condescendiente, pero en modo alguno se trataría de una racionalidad de segundo grado merecedora de especial estima. **La voluntad de creer de la autotrascendencia tiene que estar acompañada de una voluntad de racionalidad, es decir, del deseo de tener sólo creencias para las que haya buenas razones.**

5. La paradoja de la autotrascendencia

Llegamos aquí a la principal dificultad que plantea la autotrascendencia: que en ella se reproduce, a otro nivel, la paradoja de la irracionalidad. **La autotrascencencia implica que hay una creencia que el sujeto desea tener, y este deseo tiene que estar de alguna manera centrado en la verdad. Esto plantea la pregunta de qué razones puede tener el sujeto para desear tener la creencia.**[7] **Si el**

[7] Dos respuestas típicas a esta pregunta son las de que la creencia en cuestión te hace más feliz, o te incita a llevar una vida más decente. Pienso que estos planteamien-

sujeto carece de razones que justifiquen la creencia, su comportamiento sería, no un ejercicio de metarracionalidad, sino de metairracionalidad: se trataría del intento de autoinducirse causalmente, y no por razones, una creencia para la que no se tienen razones (sería un caso de autocorrupción). Pero si el sujeto sí tiene razones que justifican la creencia, entonces ya tiene la creencia y no necesita inducírsela. Es decir, si el sujeto no cree racionalmente que *p* pero se autoinduce no racionalmente la creencia de que *p*, estaríamos ante un autoengaño planificado; si el sujeto tiene razones para creer que *p*, no necesita autoinducirse la creencia de que *p* de un modo puramente causal, no necesita autotrascenderse.

6. Ideas y creencias

Esto sugiere que, para que tenga sentido hablar de autotrascendencia, en lugar de, o bien autoengaño planificado (en el mejor de los casos, benigno) o bien, simplemente, objetividad, tiene que haber un sentido en el que el sujeto crea lo que desea creer y un sentido en el que no lo crea. Para expresar esta diferencia entre estos dos sentidos, o dos formas de creencia, recurriré a la distinción orteguiana entre ideas y creencias.

José Ortega y Gasset contraponía lo que él llamaba creencias a lo que él llamaba ideas (o ideas-ocurrencias). Las ideas son las opiniones que tenemos o que nos formamos, el conjunto de opiniones y normas de conducta que sostenemos o defendemos. Las creencias son aquello con lo que contamos, el trasfondo de supuestos no formulados que, más que justificar, dan sentido a nuestras ideas. En palabras de Ortega, "en la creencia se está, y la ocurrencia se tiene y se sostiene. Pero es la creencia quien nos tiene y sostiene a nosotros" (Ortega, 1940, p. 384).[8]

La visión de la autotrascendencia que propongo, basada en estas nociones, es la siguiente. Un sujeto, y especialmente un sujeto racional y crítico, puede experimentar un conflicto entre sus ideas y sus creencias. Puede tener pautas de comportamiento, formular espontáneamente juicios o realizar inferencias que sólo tienen sentido si se le atribuye una

tos, o bien son imposibles por pragmáticamente paradójicos, o bien, si lo fueran, conducirían a la desintegración del mundo real y la paranoia de que habla Williams. Pero, aunque, en aras del argumento, aceptáramos su posibilidad, estos planteamientos no darían cuenta de la autotrascendencia, sino que la reducirían al autoengaño.

[8] Una noción equivalente en lo fundamental al concepto orteguiano de creencia es, en la tradición analítica, la tesis del trasfondo de John Searle (cf. p.e. el capítulo 5 de Searle, 1983). Obsérvese que el término "creencia" en Ortega es menos amplio que en Davidson; éste último comprendería tanto las creencias como las ideas en sentido orteguiano.

determinada creencia de trasfondo. Al mismo tiempo, puede estar sinceramente convencido de que esa creencia es falsa, y tener para ello razones que considera buenas. En el lenguaje orteguiano, esta persona viviría en una creencia que no sostiene.

Imaginemos, por ejemplo, que un sujeto tiende a adoptar una actitud de desconfianza ante los miembros de un determinado grupo social y que esa actitud sólo se explica si se le atribuye la creencia de que los miembros de ese grupo son personas en general deshonestas. No obstante, este mismo sujeto puede tener la opinión de que las personas de ese grupo no son más deshonestas que el resto. Por ejemplo, si se le preguntara si opina que las personas de ese grupo son más deshonestas que la media, contestaría sinceramente que no, y además estaría en condiciones de producir buenos argumentos en defensa de su respuesta, de dar razones a la vez buenas y sinceras que la justifican.

Esta posible situación, que es cualquier cosa menos rara, no es un caso de hipocresía (suponemos que el sujeto es sincero cuando afirma que no opina que los miembros de ese grupo son deshonestos). Pero tampoco es sin más una contradicción entre dos ideas, puesto que es demasiado flagrante para ser sostenida por un sujeto mínimamente racional (y suponemos que el sujeto en cuestión lo es). La manera de hacer inteligible esta situación tan familiar es, a la manera davidsoniana, atribuir al sujeto una mente dividida. Pero esta división no se produce entre, por así decirlo, dos compartimentos de sus ideas, sino entre una parte de sus creencias y algunas de sus ideas.

En este punto, hay dos situaciones posibles. La primera es que el sujeto permanezca inconsciente de la contradicción entre sus ideas y sus creencias. La segunda, que es la que nos interesa, es que el sujeto perciba esta contradicción. En este segundo caso, salvo que se trate de un cínico (y supondremos también que no lo es), el sujeto se formará una mala opinión sobre su creencia, que percibirá como un prejuicio que no desea tener y que además da lugar a comportamientos suyos con los que no se siente moralmente reconciliado. El problema es que al sujeto no le basta con autoconvencerse de que los miembros de ese grupo social no son particularmente deshonestos. Si eso bastara, el problema no se daría, porque en realidad ya está convencido de ello. No se trata pues de elaborar un buen razonamiento que refute su prejuicio, porque, para el sujeto, el prejuicio está ya refutado.

Tenemos, por lo tanto, una creencia activa, que influye en los juicios, razonamientos y comportamientos del sujeto, pero que es contradictoria con ideas suyas basadas en buenas razones. La salida de esta situación, en consecuencia, exige el tipo de reacción que hemos denomi-

nado autotrascendencia. Es decir, no el autoconvencimiento de la falsedad de la creencia por medio de razones (el sujeto ya está convencido de ello), sino la modificación de sus creencias tratando de provocar estados psicológicos que tiendan a causar la creencia contraria. O, más exactamente, que tiendan a convertir en creencia lo que para él es sólo una idea inerte. En el caso que nos ocupa, son muchas las maneras en que el sujeto puede hacer esto. Puede, por ejemplo, frecuentar más a miembros de ese grupo social o intimar más con los que ya frecuenta; evitar la compañía de personas hostiles a ese grupo; implicarse en organizaciones dedicadas a combatir ese tipo de discriminación y escuchar el testimonio de sus víctimas; leer literatura o escuchar música de autores de ese grupo social, etc. Nada de ello, repito, le va a convencer de nada de lo que no esté convencido, pero sí hará que ese convencimiento deje de ser una opinión verdadera y racional pero inerte y pase a ser una creencia arraigada y activa, en lenguaje orteguiano, una creencia en la que vive y no sólo una idea que sostiene.

Imaginemos ahora otro caso. Supongamos que una persona está absolutamente convencida, y por razones que le parecen buenas, de que está en la obligación moral de contribuir al exterminio de un grupo social. Por otro lado, siente una repugnancia instintiva ante semejante acción, a la que no puede evitar ver como una salvajada. Ante este dilema, opta por dejarse llevar por su rechazo a matar a esas personas y se abstiene de colaborar en el exterminio. Con ello, contradice unas ideas según las cuales el exterminio es una obligación moral, **a pesar de que esas ideas son las suyas**. El candidato a genocida no ha refutado con razones de ninguna clase la opinión de que su deber es colaborar en el exterminio. Sí ha decidido dejarla en suspenso, convertirla en una idea inerte. Para superar esta escisión entre sus creencias y sus ideas, el sujeto puede intentar provocar estados psicológicos que tiendan a causar el distanciamiento, la duda, el escepticismo o el rechazo a esas ideas. Por ejemplo, frecuenta personas con ideas opuestas a las suyas, lee libros y periódicos contrarios a ellas, ve películas inspiradas en ideas contrarias, fija su atención en rasgos negativos de la gente que defiende sus ideas, etc. Es decir, se somete a sí mismo a una especie de lavado de cerebro. La conciliación entre ideas y creencias, si finalmente se produce, será el resultado de la adopción de ideas diferentes basadas en razones. Pero la decisión de considerar las propias ideas "refutadas" por intuiciones no teóricas ni razonadas es anterior y, habitualmente, condición de posibilidad de la búsqueda de razones contra ellas.

En las dos situaciones descritas, el sujeto se autotrasciende, es decir, emprende el tipo de autocrítica y autosuperación que Davidson considera "la esencia misma de la racionalidad y la fuente de la libertad". Pero lo hace en direcciones opuestas. El primero se esfuerza por provocar cau-

salmente estados psicológicos que debiliten unas creencias que le resultan indeseables y le induzcan otras creencias que se adecuen a sus ideas. En el segundo caso, el sujeto se siente incómodo con sus ideas, precisamente porque contradicen sus creencias. En este caso, la percepción del conflicto adopta la forma de la desconfianza o incluso temor hacia algunas de las ideas y los argumentos en los que éstas se basan, a pesar de que se piense que las opiniones son verdaderas y las razones que la apoyan son buenas. La autotrascendencia aquí consistiría en un "instalarse" voluntariamente en las propias creencias, es decir, en evitar alterar el tipo de juicios, razonamientos y comportamientos que el sujeto tiende a realizar espontánea e irreflexivamente, al tiempo que se decide "poner entre paréntesis" las ideas que las contradicen.

7. Conclusión

La visión de la autotrascendencia que he esbozado se basa, por un lado, en la idea davidsoniana de la división de la mente y, por otro, en la dimensión experiencial de las creencias que he tratado de ilustrar apoyándome en la distinción orteguiana entre ideas y creencias. Esta visión nos permite dar un sentido coherente a la idea de una intervención de la voluntad en la formación de las creencias (en el sentido amplio, no en el orteguiano) que no se reduzca a la búsqueda de la objetividad y que pueda ser reivindicada como racional, e incluso como particularmente valiosa.

El tipo de intervención de la voluntad que hemos venido llamando autotrascendencia es lo que nos permite evitar instalarnos en creencias arraigadas no basadas en razones. Es decir, se trata de la decisión de no adoptar como razones lo que no son sino excusas (el "me educaron así", "tuve tales experiencias que me marcaron", "soy así"). En sentido opuesto, es lo que nos permite evitar que nuestra humanidad sea secuestrada por doctrinas dogmáticas. En última instancia, lo que subyace a la autotrascendencia es la autopercepción como un sujeto dividido y el intento, necesariamente incompleto, de ser un sujeto único. Lo que implica que, en nuestro conocimiento del mundo, el autoconocimiento desempeña un papel esencial.

No obstante, la autotrascendencia tiene un "lado oscuro": lo que, con Davidson, hemos llamado autocorrupción. Cada conflicto entre ideas y creencias puede intentar resolverse de dos maneras: tratando de adecuar las ideas a las creencias o, al contrario, tratando de adecuar las creencias a las ideas. Los dos ejemplos que he puesto son susceptibles de la "solución" opuesta. El genocida potencial puede tratar de superar sus escrúpulos espontáneos a base de imbuirse más de las ideas que le llevan a cometer el crimen y tratar de insensibilizarse ante el sufrimiento

de sus víctimas, de verlas como seres infrahumanos, como culpables que reciben un castigo proporcionado o como responsables de su propia situación. La persona prejuiciosa puede optar por instalarse en sus creencias prejuiciosas y tratar de desprenderse de sus ideas igualitarias.

En consecuencia, la capacidad de tener deseos de segundo orden sobre nuestros deseos, gustos, hábitos o creencias es lo que hace posibles tanto la autocrítica y la autosuperación como el dogmatismo y la autodegradación. Lo que nos lleva a la inquietante conclusión de que la condición de posibilidad de la racionalidad y la libertad es la misma que la de la irracionalidad y la alienación, como también lo son los mecanismos de unas y otras.

El único indicio fiable que se me ocurre para distinguir la autotrascendencia frente a la autocorrupción es fenomenológico: la autotrascendencia, al contrario que la autocorrupción, sencillamente, no resulta agradable, no genera, de manera inmediata, ni bienestar ni buena conciencia. Como dice Nietzsche:

> ¿Es que la salvación, la bienaventuranza, o, por decirlo con una palabra más técnica, el *placer* constituye una prueba de la verdad? Casi se podría decir que lo que prueba es lo contrario, pues cuando nos preguntamos "¿qué es la verdad?", interviniendo en nuestro interrogante sentimientos de placer, nos sentimos inducidos a sospechar grandemente respecto a la "verdad". La prueba del "placer" es una prueba de placer, una prueba agradable, y nada más. ¿En base a qué cabe establecer que los juicios verdaderos, por el hecho de serlo, producen más placer que los falsos y que, en virtud de una armonía preestablecida, reportan necesariamente sentimientos agradables?
>
> La experiencia de todos los espíritus profundamente serios enseña *lo contrario*. Cualquier avance en el camino de la verdad se ha tenido que llevar a cabo mediante una lucha, en la que ha habido que entregar casi todo aquello a lo que se adhiere nuestra vida. Para ello, se necesita una grandeza de alma, ya que servir a la verdad constituye el más duro de los servicios. Porque ¿qué significa ser honrado en las cosas del espíritu? Significa ser inflexible con nuestro propio corazón, desdeñar los "bellos sentimientos", convertir en un caso de conciencia toda afirmación y toda negación. La fe salva, *luego* es falsa. (Nietzsche, *El Anticristo*, § 50)[9]

[9] Este trabajo se enmarca en el proyecto de investigación "El papel de la voluntad en las creencias y juicios de gusto", financiado por el Ministerio de Educación y Ciencia de España (BFF2003-08335-C03-02). Versiones previas fueron leídas, en noviem-

Bibliografía

Cavell, M., 1999: "Reason and the Gardener", en Hahn, L.E. (ed.), *The Philosophy of Donald Davidson*, Open Court, Chicago y La Salle.

Davidson, D., 1982: "Paradoxes of Irrationalty", en Wollheim y Hopkins (eds.), *Philosophical Essays on Freud*, Cambridge U.P.; reimp. en Davidson, 2004.

——— 1985: "Incoherence and Irrationality", *Dialectica*, 39; reimp. en Davidson, 2004.

——— 1986: "Deception and Division", en Elster (ed.), *The Multiple Self*, Cambridge U.P.; reimp. en Davidson, 2004.

——— 1997: "Who is Fooled?", en Dupuy (ed.), *Self-Deception and Paradoxes of Irrationality*, CSLI, Stanford; reimp. en Davidson, 2004.

——— 2004: *Problems of Rationality*, Clarendon Press, Oxford.

Nietzsche, F., *El Anticristo (Maldición sobre el cristianismo)*, Alianza, Madrid, 2001.

Ortega y Gasset, J., 1940: "Ideas y creencias", en *Obras Completas V*, Revista de Occidente, Madrid, 1947.

Searle, J., 1983: *Intentionality. An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press.

Williams, B., 1970: "Deciding to Believe", en Kiefer y Munitz (eds.), *Language, Beliefs and Metaphysics*, State of New York Press, Albany; reimp. en Williams, B., *Problems of the Self, Philosophical Papers 1956-1972*, Cambridge U.P., 1973.

Campus de Espinardo
30071-Murcia (España)
Teléfono: 34 968 363466
Fax: 34 968 364266
mhi@um.es

bre de 2005, en el Coloquio en homenaje a Donald Davidson celebrado en Montevideo (organizado por Carlos Caorsi) y en la Sociedad Argentina de Análisis Filosófico (Buenos Aires) y, en enero de 2006, en la Universidad Pompeu Fabra (Barcelona). El texto final debe mucho a las observaciones de los participantes en estos tres actos.