

# Autoengaño y voluntarismo doxástico \*

## Self-deception and doxastic voluntarism

Por: **Gustavo Fernández Acevedo**

Departamento de Filosofía  
Facultad de Humanidades  
Universidad Nacional de Mar del Plata  
Mar del Plata, Argentina  
E-mail: facevedo@mdp.edu.ar

Fecha de recepción: 9 de noviembre de 2017  
Fecha de aprobación: 12 de diciembre de 2017  
Doi: 10.17533/udea.ef.n57a07

**Resumen.** *La doctrina conocida como “voluntarismo doxástico”, esto es, la idea según la cual podemos adoptar libremente nuestras creencias, ha sido considerada insostenible por la mayoría de los filósofos. Ahora bien, cabe distinguir dos versiones de esta doctrina. El voluntarismo doxástico directo, que coincide con la descripción precedente, y el voluntarismo doxástico indirecto, que sostiene la tesis más moderada según la cual sería posible inducir en nosotros mismos determinadas creencias de manera indirecta por medio de la modificación planificada de nuestras conductas. Bernard Williams, quien ha presentado influyentes argumentos en contra del voluntarismo doxástico directo, ha admitido que su variante indirecta es posible, pero sólo por medio del autoengaño. En este artículo se examinan distintas concepciones contemporáneas respecto del autoengaño y se concluye que bajo ninguna de las caracterizaciones presentadas este fenómeno puede constituir un medio eficiente para la inducción voluntaria indirecta de creencias.*

**Palabras clave:** *Autoengaño, voluntarismo doxástico, conducta, acción, creencias, voluntad*

**Abstract.** *The doctrine known as “doxastic voluntarism”, that is, the idea according to which we can freely adopt our beliefs, has been considered as untenable by most philosophers. Moreover, it is possible to distinguish two versions of this doctrine. The direct doxastic voluntarism corresponding to the latter description; and the indirect doxastic voluntarism, that sustains a more moderate thesis according to which it would be possible to induce certain beliefs in ourselves in an indirect way through the planned modification of our conducts. Bernard Williams, who has presented influential arguments against the direct doxastic voluntarism, has admitted that its indirect variation is possible, but only via self-deception. In this article, different contemporary conceptions of self-deception are examined, to conclude that under any of these can this phenomenon constitute an efficient means for the indirect voluntary induction of beliefs.*

**Keywords:** *Self-deception, doxastic voluntarism, conduct, action, beliefs, will*

---

\* El artículo es resultado del proyecto de investigación titulado *El rol de la evidencia en el autoengaño*. Inicio: 01/01/2017; fin: 31/12/2017. Entidad que financió el proyecto: Universidad Nacional de Mar del Plata. Deseo agradecer las observaciones de los árbitros de *Estudios de Filosofía*, que me permitieron mejorar varios pasajes de la versión previa de este trabajo.

### Cómo citar este artículo

MLA: Fernández, Gustavo. “Autoengaño y voluntarismo doxástico.” *Estudios de Filosofía*, 57 (2018): 139–160

APA: Fernández, G. (2018). Autoengaño y voluntarismo doxástico. *Estudios de Filosofía*, 57, 139–160.

Chicago: Fernández, Gustavo. “Autoengaño y voluntarismo doxástico”. *Estudios de Filosofía* n.º 57 (2018): 139–160.

## I

La idea de que podemos ejercer un control voluntario sobre aquello que creemos, esto es, que podemos decidir adoptar libremente nuestras creencias, constituye la posición comúnmente conocida como *voluntarismo doxástico*. De acuerdo con esta tesis, así como algunas de nuestras acciones (por ejemplo, levantar un brazo) son acciones básicas, que no requieren para su ejecución de la realización de ninguna acción adicional, ciertos casos de adquisición de creencias constituirían acciones básicas: podríamos adoptar algunas creencias de modo voluntario y directo, sin vernos forzados a ello por la particular relación entre las creencias y la realidad. La defensa del voluntarismo doxástico no requiere afirmar que *cualquier* creencia puede ser candidata a la adquisición directa; sólo se necesita sostener que *algunas* creencias pueden ser adquiridas de ese modo. Tal presunta conexión entre acciones y creencias ha inspirado la distinción entre diferentes tipos de voluntarismo doxástico. Así, Nottelman (2006) presenta, siguiendo a Audi, una división de esta doctrina en dos versiones: versión conductual y versión genética. La primera, más fuerte y menos defendida, sostiene que creer constituye un tipo de acción, algunos de cuyos casos son directamente voluntarios; la segunda, más débil, expresa el punto de vista según el cual la formación de una creencia es en ocasiones una clase tal de acción.

En cualquier caso, y ya desde Hume, ha existido un amplio consenso en que el voluntarismo doxástico es insostenible. Distintos filósofos han afirmado que es completamente imposible elegir a voluntad y de modo directo nuestras creencias (Adler, 2002; Elster, 1979; Williams, 1973). Para los críticos, un rasgo esencial de las creencias es que tienden o apuntan hacia la verdad: se imponen por el peso de la evidencia en su favor, y no por nuestro deseo de que el estado al que hacen referencia acontezca; esta tesis se ha denominado, consecuentemente, *involuntarismo doxástico*. Si fuese posible para nosotros creer a voluntad a sabiendas, se afirma, sería posible creer que *p* es verdadero, sabiendo todo el tiempo que la creencia no está justificada por la evidencia, que fue adquirida a voluntad y que las creencias pueden ser adoptadas sin consideración a su verdad o falsedad, todo lo cual es considerado imposible. Esta imposibilidad puede ser concebida como una imposibilidad conceptual/lógica/metafísica o como una imposibilidad psicológica y, aun cuando fuesen superables las objeciones relativas al primer tipo de imposibilidad, esto no implicaría que también lo serían las críticas referentes al segundo tipo. Se han presentado, no obstante, diversas réplicas a los argumentos contra el voluntarismo doxástico (Bennett, 1990; Funkhouser, 2003) e intentos de

mostrar como las dificultades de esta posición pueden ser superadas (Ryan, 2003; Steup, 2008).

A los fines del presente trabajo, resulta necesario distinguir entre dos tipos de voluntarismo doxástico habitualmente diferenciados: el voluntarismo doxástico *directo* (que se ajusta a la noción inicialmente descrita) y el voluntarismo doxástico *indirecto*. Según esta última variante sería posible, en principio, inducir en nosotros mismos determinadas creencias de manera indirecta por medio de la modificación planificada de nuestras conductas; esta alternativa suele ser ejemplificada por el consejo de Pascal de actuar como si se fuese creyente, de modo de inculcar en nosotros mismos la creencia en Dios. La posibilidad del voluntarismo doxástico indirecto es admitida por los críticos más firmes de la variante directa de esta tesis. Adler (2002) observa que, en lo que respecta al voluntarismo doxástico, la pregunta pertinente no es si somos capaces de influir o controlar indirectamente nuestras creencias; en su opinión, es innegable que podemos inducir en nosotros mismos una creencia prestando atención a fuentes que expresen un único punto de vista, sesgándonos hacia la opinión buscada. El aspecto positivo de esta capacidad, agrega, es que podemos desarrollar disposiciones o hábitos para estar abiertos a la crítica o a la evidencia contraria, y evaluarlas equitativamente. Bernard Williams (1973), por su parte, pese a haber desarrollado un argumento sumamente influyente en favor de la imposibilidad conceptual del voluntarismo directo, concede que hay lugar para aplicar la decisión de creer por medio de rodeos [*roundabout*]. No obstante, considera que esta alternativa es profundamente irracional y que puede tener éxito sólo por medio del autoengaño.

Ahora bien, el esclarecimiento de la naturaleza de los vínculos entre el autoengaño y la inducción voluntaria indirecta de creencias requiere de un análisis mayor al que Williams lleva a cabo. De hecho, este autor señala explícitamente que no se ocupará del autoengaño en su trabajo y, más allá de algunas observaciones relativas a las implicaciones de este fenómeno para el manejo de la evidencia, no hay intento alguno de dilucidar su naturaleza. Sin embargo, la necesidad de un mayor análisis no se debe sólo a la ausencia de un examen de esta cuestión por parte de Williams, sino principalmente a la falta de acuerdo respecto de cuáles son los estados que originan y hacen posible el autoengaño y, por ende, a qué clase de relación puede tener con el voluntarismo doxástico indirecto. En consecuencia, la determinación de los vínculos entre ambos fenómenos constituye una tarea pendiente, que afrontaremos en el presente trabajo.

El artículo estará organizado de la siguiente forma. En la segunda sección examinaremos algunas posibles conexiones entre el voluntarismo doxástico indirecto y el autoengaño, y nos detendremos especialmente en un intento de mostrar que es posible inducir conscientemente en nosotros mismos creencias que consideramos falsas de un modo indirecto sin que el autoengaño se encuentre involucrado; sostendremos, respecto de esta tentativa, que hay varias razones para pensar que no logra su objetivo. En la tercera sección presentaremos algunas notas en principio imprescindibles del autoengaño, centrándonos en una dimensión particular, esto es, la referente a los estados mentales que le dan origen. En la última sección examinaremos los problemas de las dos teorías más influyentes acerca de los procesos que originan el autoengaño, esto es, el intencionalismo y el no intencionalismo, y analizaremos su posible rol en el intento de inducir voluntariamente creencias en nosotros mismos; sostendremos que, cualquiera que fuere la concepción que se adopte respecto de los estados mentales que originan el autoengaño, este fenómeno no puede constituir un medio eficiente para la inducción voluntaria de creencias.

## II

Como veremos en las siguientes secciones, suele atribuirse un carácter indirecto al proceso que conduce a un estado de autoengaño. Sin embargo, y pese a que el voluntarismo doxástico directo no parece requerir de ninguna mediación, no siempre se ha dado por sentado a cuál variante de este fenómeno puede supuestamente prestar apoyo. Booth (2007) examina la posibilidad de que el autoengaño constituya una vía de justificación para el defensor del voluntarismo doxástico directo, que permitiría eludir los argumentos a favor de la imposibilidad psicológica (suponiendo, claro está, que la imposibilidad conceptual/lógica/metafísica haya sido superada). Esta posibilidad plantea el problema del modo en que el autoengaño debe ser concebido para lograr ese objetivo. Booth considera que, con ese fin, el fenómeno debe ser interpretado de modo tal que a) muestre cómo al menos algunos casos de autoengaño son logrados sin la intervención de ninguna mediación, y b) que tales actos sean conducidos deliberadamente y con plena conciencia. A partir del análisis del clásico ejemplo de mala fe presentado por Sartre (1943), el de la joven que se niega a reconocer ante sí misma los intentos de seducción de su acompañante, Booth considera que no es posible satisfacer ambas condiciones simultáneamente: no es posible cumplir con la condición a) sin violar la condición b), y a la inversa; consecuentemente, dado que ambas deben ser cumplidas para que los casos de autoengaño constituyan casos de voluntarismo doxástico directo, parecería que

ningún acto de autoengaño podría establecer la posibilidad psicológica de control doxástico directo. El autoengaño puede, concluye Booth, involucrar un control doxástico indirecto, pero no un control directo.

Si Booth estuviera en lo correcto (cuestión que no examinaremos aquí) entonces el rol del autoengaño en el voluntarismo doxástico quedaría, en el mejor de los casos, acotado a la variante *indirecta* de este fenómeno. Como hemos señalado, esta posibilidad ha sido admitida por filósofos como Williams, quien considera que el autoengaño constituye una condición necesaria para el voluntarismo doxástico indirecto.<sup>1</sup> Sin embargo, hay autores que han sostenido que es posible decidir creer

---

1 Uno de los evaluadores de este artículo ha señalado que puede darse una interpretación a las observaciones de Williams que difiere con aquella a la que aquí adhiero, esto es, que el autoengaño constituye una condición necesaria para la inducción voluntaria indirecta de creencias. El árbitro señala acertadamente, entiendo, que el texto de Williams no es explícito respecto de los vínculos entre ambos fenómenos, y que algunas de sus afirmaciones pueden dar lugar a una interpretación alternativa, que consiste en sostener que este autor en realidad rechaza la idea de que los casos en los que por medio del autoengaño llegamos a creer algo en lo que deseamos creer constituyen casos genuinos de inducción voluntaria indirecta de creencias. El árbitro considera que, si tal interpretación alternativa fuese correcta, el artículo no habría identificado correctamente el foco de discusión. Al respecto corresponde efectuar varias observaciones. En primer lugar, creo que la interpretación alternativa sugerida por el árbitro, si bien plausible, lo es menos que aquella a la que adhiero en este artículo. Si lo entiendo correctamente, el argumento principal del evaluador se basa en que de la afirmación de Williams por la cual rechaza “el *proyecto consciente* para hacerme creer lo que deseo creer” (cursivas del árbitro) parece seguirse que el autoengaño nos lleva a creer aquello que deseamos justamente porque opera de modo inconsciente. Pero si se supone que la inducción voluntaria indirecta de creencias requiere de algún grado de consciencia en al menos ciertos momentos del proceso, el que Williams parezca sostener que el autoengaño opera de modo inconsciente constituiría una razón para pensar que no considera que los casos en los cuales logramos creer algo por medio del autoengaño no son casos reales de inducción voluntaria indirecta de creencias. Sin embargo, la discusión de Williams acerca de las posibles objeciones al intento de inducir en uno mismo aquello que sabemos que es falso indicaría otra cosa. Estas objeciones apuntan a que tal intento implicaría o bien un proceso de remoción de toda evidencia contraria a la creencia deseada, lo que en última instancia conduciría “a la paranoia” (151), o bien, si queremos evitar esta consecuencia, a conducirnos en el mundo de modo tal de evitar cualquier evidencia que comprometa nuestra creencia. Pero, si esto último fuera el caso, agrega, este proyecto de autoengaño requeriría que quien lo practica realmente conozca la verdad, esto es, sepa que la creencia que favorece es falsa; si esto no fuese así, señala, no sería capaz de conducirse de modo tal de evitar la evidencia contraria. Esto último, considero, implica que quien se autoengaña es, en alguna medida, consciente de la verdad, por lo que el autoengaño no posee el carácter plenamente inconsciente que la interpretación del árbitro le atribuye. Y, en consecuencia, la incompatibilidad entre la inducción voluntaria indirecta de creencias y el autoengaño no sería tal. Admito que, como señalé, el análisis de Williams plantea cuestiones espinosas y su posición no es explícita sobre varias de ellas, por lo cual resulta difícil encontrar un argumento que zanje la discrepancia. No obstante, la posibilidad de que mi interpretación sea errónea no elimina el interés de la tesis defendida en el presente artículo; entiendo que la necesidad de establecer con mayor precisión los vínculos entre voluntarismo doxástico y autoengaño no se deben solo (ni principalmente) al

de modo indirecto sin que el autoengaño esté involucrado. Este es el caso de Thomas Cook (1987), cuya defensa de esta tesis examinaremos a continuación.

Cook observa que diversos filósofos, entre los cuales se cuentan Elster, Pears y el citado Williams, sostienen que el esfuerzo de inducir en uno mismo la creencia en la verdad de una proposición que se considera falsa (o en una proposición para la cual se carece de evidencia que justifique su aceptación) sólo puede tener éxito si en alguna parte del proceso se olvida que se está haciendo tal esfuerzo. Aunque este punto de vista tiene un atractivo intuitivo fuerte, Cook afirma que es falso y que esta falsedad puede ser puesta de manifiesto mediante un ejemplo. Considera que la posición de esos filósofos proviene, en primer lugar, de una subestimación de los factores no probatorios que pueden influir en las creencias y, en segundo lugar, de un descuido del hecho de que los propios criterios para la justificación son en sí mismos creencias susceptibles de manipulación de forma indirecta.

Cook propone un ejemplo que muestra, a su entender, la posibilidad de creer que  $p$  mientras se piensa que tal creencia se deriva de la decisión de creer que  $p$ . Nick es un brillante y ambicioso biólogo que trabaja en una pequeña universidad fundamentalista. Tiene importantes aspiraciones de éxito profesional, y está dispuesto a hacer grandes esfuerzos por lograrlo. Sin embargo, advierte correctamente que sus ideas creacionistas reducen a un mínimo sus posibilidades de obtener tal éxito. Ha considerado la posibilidad de mantener ocultas sus creencias creacionistas mientras acepta públicamente la teoría de la evolución, pero una vida de hipocresía y engaño no es un porvenir que le resulte atractivo. Así, estima que la mejor solución para su problema consiste en creer que la teoría de la evolución es la concepción correcta acerca del desarrollo de la vida sobre la Tierra, de modo que decide adoptarla. Nick cree (como siempre ha creído) que la revelación expuesta en las Escrituras, interpretadas literalmente, es el árbitro último en tales disputas, de modo que será difícil para él inducir la creencia en cuestión. No obstante, Nick tiene conocimiento acerca de los factores no probatorios que influyen en la adopción de creencias: los mecanismos de reducción de la disonancia que tienden a mantener las creencias en consonancia con la conducta; las influencias sobre las creencias por parte de personas que admira y con las que comparte su tiempo; la hermenéutica y la filosofía de la ciencia, relativas a los modos en los cuales la evidencia es interpretada y qué debe ser considerado evidencia; por último, sabe que la atención selectiva a la

---

interés por elucidar la posición de Williams. Aunque este autor nunca hubiera mencionado al autoengaño como estrategia para la inducción voluntaria indirecta de creencias, la relación entre ambos fenómenos mantendría su interés teórico como casos paradigmáticos de irracionalidad motivada en el plano teórico.

evidencia y el empleo de hipótesis interpretativas son ampliamente una cuestión de hábitos. Con este conocimiento en mente, inicia el proyecto de inducir las creencias buscadas y concurre a Harvard para realizar sus estudios de posgrado. Aunque está convencido de que la teoría de la evolución es falsa, está decidido a creer que es verdadera, y confía en que el prestigio de sus instructores, las presiones sociales y la disonancia interna provocada por la necesidad constante de reconocer y utilizar las hipótesis evolucionistas (en clase, en los exámenes, etcétera) inducirá en él, en última instancia, las creencias deseadas.

De acuerdo con Williams y Elster, prosigue Cook, Nick no podrá tener éxito en su proyecto a menos que manobre de modo tal de olvidar o engañarse a sí mismo acerca de sus motivos y acerca de los orígenes de la creencia que desea adquirir. Bajo este punto de vista, en la medida en que recuerde que su creencia es una consecuencia del empleo de medios indirectos, que descansan en mecanismos e influencias no probatorios, no será capaz de creer realmente en la verdad de la teoría de la evolución. Sin embargo, Cook considera que la continuación del ejemplo permite ver la falsedad de la tesis de Williams y Elster: seis años más tarde, Nick se doctora aceptando la teoría biológica contemporánea. Recuerda sus inicios y cómo esperaba ser seducido por la creencia en tal teoría; se ríe para sus adentros de lo ignorante que era antes de haber ingresado a Harvard. Es plenamente consciente de que su creencia es consecuencia de una estrategia; no ha olvidado nada y no tiene necesidad de engañarse a sí mismo respecto de nada. Se podría argumentar, observa Cook, que Nick ha olvidado que las Escrituras son el árbitro final de las cuestiones en disputa; sin embargo, no lo ha olvidado, simplemente sostiene ahora que esa creencia es falsa y que él estaba en un error al sostener que era verdadera. Quizás, entonces, ha olvidado la evidencia que una vez le condujo a sostener sus creencias previas. Pero, de hecho, no lo ha olvidado en absoluto; se trata solamente de que ahora está inclinado a pensar que esa evidencia era muy débil y que sus creencias previas carecían mayormente de apoyo probatorio, y descansaban casi enteramente en la ignorancia, en presiones sociales, en la superstición y en la aceptación injustificada de las concepciones de sus padres y maestros fundamentalistas. La historia que cuenta ahora Nick acerca de su posición epistémica previa es una redescipción de esa posición, pero tal redescipción no requiere de olvido selectivo o autoengaño.

La mayoría de nosotros, dice Cook, estaría de acuerdo en que Nick pasó de una creencia que cuenta con poca evidencia en su apoyo a una creencia que cuenta con mayor evidencia favorable. Pero, agrega, no debería pensarse que nuestro juicio acerca de la fuerza del ejemplo depende de este hecho. Si el caso hubiera sido a la



inversa, Nick, un brillante evolucionista que quiere creer en la teoría creacionista, tampoco hubiera sido autoengañado. Nuestro criterio epistémico puede ser alterado por numerosas vías no racionales. En la medida en que uno se las arregle para modificar tal criterio, concluye, las propias creencias pueden ser conscientemente manipuladas sin necesidad de autoengaño.

¿Ha probado Cook que pueden existir casos de voluntarismo doxástico indirecto sin autoengaño? Sin duda, puede decirse mucho acerca de la plausibilidad psicológica de su ejemplo. Entre las posibles objeciones, cabe señalar que parece sumamente improbable que alguien que posee una convicción fuerte respecto de un determinado sistema de creencias, sistema que posee una importancia central en su vida (como sin duda sería el caso tratándose de un credo religioso fundamentalista adquirido y mantenido desde la infancia), esté simplemente dispuesto a desecharlo motivado por un interés como el éxito profesional, cuya relevancia psicológica parece comparativamente mucho menor. Aun en caso de que esto pudiese intentarse, parece más improbable aun que pueda hacerse el intento sin un costo psíquico sumamente elevado, costo que, podemos suponer, pondría en riesgo permanente el éxito de la empresa. Sin embargo, las dudas van más allá de la implausibilidad psicológica del caso propuesto, ya que se extienden a la interpretación según la cual se trata de un caso de voluntarismo doxástico indirecto. Una crítica de las conclusiones expuestas por Cook podría consistir en una defensa de alguna de las siguientes opciones: o bien hay voluntarismo doxástico indirecto pero, contra su tesis, mediado por el autoengaño, o bien no hay voluntarismo doxástico, caso en el que es irrelevante el que haya o no autoengaño. Examinaremos esta última opción en primer lugar.

Santamaría Navarro (2009), en referencia a las afirmaciones de Cook relativas al caso inverso (esto es, de la defensa de la verdad del evolucionismo a sostener la verdad del creacionismo) observa que si la tesis de este autor fuese suficientemente fuerte podría haber elegido tal alternativa; sin embargo, no fue ésta la opción elegida. Considera que el caso propuesto por Cook, claramente, no constituye un caso de creencia a voluntad. Parecería más bien que Nick desea creer en algo que no cree y, en el intento de lograrlo, descubre que la evidencia apoya la creencia deseada.<sup>2</sup>

---

2 No carece de interés esbozar una posible interpretación del caso de Nick según la cual no existiría autoengaño en su intento de adoptar la teoría de la evolución. Esta interpretación estaría basada en las condiciones sugeridas por Mele (2001) para la adopción autoengañoso de una creencia, que requieren que tal adopción no tenga lugar a través de una cadena causal desviada. El concepto de cadena causal



El punto que aquel autor quiere sostener es que lo que convence a Nick no es su deseo de creer, sino la evidencia. Por supuesto, agrega, si no hubiera deseado creer quizás no habría encontrado esta evidencia, pero tal caso le parece cercano a ciertos ejemplos propuestos por James o Russell en los que se actúa a partir de una hipótesis, y se encuentra evidencia en el curso de este proceso. La observación de Santamaría Navarro, en caso de ser correcta, constituye una forma de refutar la posición de Cook: en el ejemplo planteado puede haber (o no) autoengaño, pero en cualquier caso no habría voluntarismo doxástico indirecto; en tal caso, Cook no ha probado que puede existir voluntarismo doxástico indirecto *sin* autoengaño.<sup>3</sup> Si la objeción planteada por Santamaría Navarro pudiese generalizarse a cualquier caso de autoinducción de creencias verdaderas, entonces no podría existir voluntarismo doxástico indirecto sin autoengaño de la forma que Cook sugiere, pero simplemente porque en tales casos no habría voluntarismo doxástico indirecto.<sup>4</sup>

---

desviada suele ser presentado mediante ejemplos como el siguiente: supóngase que, en un concierto, un espectador pretende alterar al director de la orquesta, con quien mantiene una antigua enemistad personal. Para lograr esto tiene la intención de toser ruidosamente durante el concierto, pero esta misma intención causa en él un estado de nerviosismo que lo hace toser de modo ruidoso. El espectador tenía la intención de toser y esa intención causó que tosiera, pero esto último no ocurrió de modo intencional. Siguiendo este modelo, y teniendo en cuenta la interpretación de Santamaría Navarro, el ejemplo de Nick podría ser considerado un caso de cadena causal desviada. Esto es, la acción intencional de Nick consistente en ingresar a Harvard inicia una cadena causal de sucesos que finaliza en el estado buscado (la adopción de la teoría evolucionista), pero no del modo en que él lo planificó, esto es, mediante la actuación de los factores no probatorios que influyen en la adopción de creencias; lo que ha ocurrido es que, en este proceso, Nick ha encontrado buenas razones epistémicas para adoptar la teoría de la evolución, y ha resuelto adherir a ella. Lo anterior no responde, por supuesto, a la pregunta de si se trata de un caso de voluntarismo doxástico indirecto, aun cuando no exista autoengaño.

- 3 En términos similares a los de Santamaría Navarro parece expresarse Rosell (2008), en el contexto de una crítica a la tesis de James según la cual en ciertas situaciones es posible creer a voluntad: “Alternativamente, también podría suceder que, mediante mecanismos indirectos de formación de creencia, el sujeto acabe adquiriendo una creencia. Sin embargo, este tipo de voluntarismo —ampliamente reconocido— no es lo que está aquí en cuestión. Una cosa es decidir creer que *p* voluntaria y directamente y otra distinta decidir inducir en mí la creencia de que *p*. Esto último no está constreñido por las necesidades psicológica y conceptual anteriores, pues decidir formarme una creencia no es sí mismo una creencia. En estos casos es posible, y —dependiendo de qué acciones concretas se lleven a cabo— muy probable, conducirse uno mismo a creer que *p*; por bien que no es una ruta completamente segura. En este sentido, el argumento de la imposibilidad conceptual de creer voluntariamente no impide la existencia de fenómenos como el pensamiento desiderativo o el autoengaño. *En todo caso, la creencia resultante no será fruto directo de nuestra voluntad, ni creeremos que p porque queríamos o estábamos empeñados en creer que p, sino porque ahora la evidencia nos dicta que p*” (204. *Cursivas nuestras*).
- 4 Obviamente, tampoco habría autoengaño: no puede existir autoengaño en caso de que alguien adopte una proposición verdadera sobre la base de la mejor evidencia disponible.

Por otra parte, la sugerencia de Santamaría Navarro de tener en cuenta el caso inverso al propuesto por Cook (esto es, Nick considera que el evolucionismo es verdadero pero desea creer en el creacionismo) pone en primer plano la cuestión relativa a la necesidad de que la creencia a adoptar sea falsa para poder realizar una atribución de autoengaño. ¿Puede haber autoengaño en caso de que la creencia (o conjunto de creencias) a adoptar sean verdaderas? Si le creemos a Mele (2001), no. Según él, una condición necesaria para que se ingrese en un estado de autoengaño consiste en que la creencia a adoptar debe ser falsa. Para sostener esto se basa en una cuestión puramente lingüística: por definición, observa, una persona es *engañada* respecto de la creencia de que  $p$  sólo si  $p$  es falsa, y lo mismo ocurre en el caso de que alguien esté autoengañado respecto de la creencia en  $p$ . En consecuencia, en caso de que Cook tuviera razón y en su ejemplo hubiera voluntarismo doxástico indirecto, entonces habría que aceptar su conclusión: pueden existir casos de voluntarismo doxástico indirecto sin autoengaño, ya que no puede existir autoengaño cuando la creencia que se adopta es verdadera. Ahora bien, no es obligatorio aceptar la posición de Mele. Lynch (2010), observa que hay dos casos distintos en los que es posible hablar de autoengaño, casos que deberían ser distinguidos: “estar autoengañado *en* creer que  $p$ ”, y “haberse engañado a uno mismo *para* creer que  $p$ ”. Sugiere que si bien es contradictorio decir que  $S$  está autoengañado al creer algo que es verdadero, no es contradictorio decir que  $S$  se engaña a sí mismo *para* creer algo que es verdadero. La idea anterior puede ser apoyada, observa Lynch, por los casos de engaño interpersonal. Supóngase que yo sé que  $p$  es verdadero y sé que  $X$  sólo creerá que es verdadero si escucha que  $Z$  considera que lo es, ya que  $Z$  es la única persona en la que  $X$  confía. En tal caso yo puedo mentirle a  $X$  diciéndole que he estado en contacto con  $Z$ , y que él ha dicho que  $p$  es verdadera. En esta situación  $X$  no está engañado en creer que  $p$ , ya que  $p$  es verdadera. Sin embargo, podemos decir que  $X$  fue engañado por mí *para que crea* que esa proposición es verdadera. Si la posición de Lynch es correcta, no podemos desestimar, sobre la base de una cuestión puramente conceptual, que alguien pueda encontrarse en un estado de autoengaño respecto de una creencia verdadera.

Podemos preguntar, entonces, si lo anterior es aplicable al caso de Nick. No es posible tener la certeza de que Nick ha adoptado la teoría de la evolución sobre la base de los buenos elementos de juicio que existen en su favor, así como tampoco hay razones concluyentes para creer que lo ha hecho sobre la base de pruebas inadecuadas o falsas. Sin embargo, parece razonable pensar que no se trata de una situación de empate entre las dos alternativas: las buenas razones para aceptar la teoría de la evolución superan con mucho (al menos en el estado actual de nuestro

conocimiento) a las posibles razones incorrectas o erróneas. Si esta suposición es correcta, entonces Nick efectivamente sería persuadido de la verdad de la teoría de la evolución, una teoría verdadera (o al menos aproximadamente verdadera), por buenas pruebas (o al menos suficientemente buenas) y, en consecuencia, no estaríamos en presencia de un caso de autoengaño. Ahora bien, en tal caso tampoco estaríamos ante un ejemplo de voluntarismo doxástico, ya que la aceptación de la teoría por parte de Nick tendría lugar de un modo racional, aun cuando sus motivos originales no buscaran la adquisición racional de la creencia deseada.

Debe resultar claro que la refutación del intento de Cook no prueba que Williams y otros estén en lo correcto al afirmar que sólo es posible el voluntarismo doxástico indirecto mediado por el autoengaño. Cook no sostiene que nunca interviene el autoengaño en la adquisición voluntaria indirecta de creencias; tampoco, entiendo, admite que el autoengaño puede ser condición suficiente para tal inducción. Simplemente afirma que existen casos en los que esta adquisición voluntaria indirecta es posible *sin* autoengaño. Si tuviera razón, el autoengaño no constituiría una *condición necesaria* para el voluntarismo doxástico indirecto; sin embargo, dadas las objeciones presentadas, el contraejemplo sugerido por él no parece lograr ese objetivo.

### III

En la sección precedente examinamos críticamente un intento de mostrar cómo es posible el voluntarismo doxástico indirecto sin la mediación del autoengaño; si bien, a los fines argumentativos, no intentamos caracterizar este último fenómeno, resulta ahora necesario detenernos en tal caracterización y en las dificultades que típicamente presenta.

El autoengaño ha constituido en las últimas décadas un problema profundamente estudiado en el ámbito de la filosofía (Demos, 1960; Paluch, 1967; Gardiner, 1970; Szabados, 1974; Bach, 1981; Davidson, 1982, 1986; Mele, 1987, 1997, 2001; Oksenberg–Rorty, 1988; Bermúdez, 2000; Funkhouser, 2005; Van Leeuwen, 2007, 2008, 2009). Atestiguan este interés los diversos estudios relativos a sus implicaciones morales, sus presuntas analogías con el engaño interpersonal, las paradojas que tal estado supuestamente implicaría (en particular, las derivadas de la exigencia de que el agente posea simultáneamente la creencia verdadera y la creencia falsa) y su hipotético carácter intencional. Más allá de las persistentes controversias respecto de su naturaleza, parece plausible suponer que una caracterización preteórica del fenómeno debería incluir las siguientes condiciones:

- a. adquisición y/o mantenimiento de una creencia falsa,
- b. frente a pruebas contrarias a ella,
- c. motivados por procesos mentales no cognitivos (por ejemplo, deseos o emociones).

Las notas precedentes, aunque mínimas, pueden ser cuestionadas por diversas razones.<sup>5</sup> La razón que más nos interesa aquí es la relacionada con un componente que, para muchos teóricos, no puede estar ausente de una caracterización adecuada del autoengaño: su presunto componente intencional. Los argumentos en favor y en contra de esta supuesta intencionalidad del autoengaño han constituido una parte fundamental de los debates sobre este fenómeno en las últimas tres décadas. En lo que resta de este apartado presentaremos brevemente las tesis definitorias de este presunto carácter intencional y también la posición alternativa, conocida como no intencionalismo o deflacionismo; examinaremos con mayor profundidad estos debates y su relación con el voluntarismo doxástico indirecto en el apartado IV.

La perspectiva según la cual el autoengaño es un fenómeno intencional y que implica la presencia de creencias contradictorias ha sido la posición estándar sobre el problema durante varias décadas. Ambas tesis derivan del supuesto de que el autoengaño debe comprenderse a la luz de los procesos de engaño interpersonal. Respecto de la primera nota, la mayoría de los teóricos (Davidson, 1986; Oksenberg–Rorty, 1988; Pears, 1985; Talbot, 1995) han coincidido en afirmar la intencionalidad del fenómeno: quien se autoengaña induce *intencionalmente* en sí mismo una creencia falsa, o al menos claramente no sustentable por las pruebas disponibles.

Donald Davidson ha sido sin duda el más influyente y destacado defensor de la tesis según la cual el autoengaño requiere de una intención para su producción:

Un agente A se autoengaña con respecto a una proposición P bajo las siguientes condiciones: A posee evidencia sobre la base de la cual cree que P es más verosímil que su negación; el pensamiento de que P, o de que debería creer racionalmente que P, ofrece a A motivos para actuar con vistas a causar en sí mismo la creencia en la negación de P (...) Todo lo que el autoengaño exige de la acción es que el motivo tenga su origen en una creencia en la verdad de P (o en el reconocimiento de que la evidencia hace más probable la verdad de P que su falsedad) y *que se lleve a cabo con la intención* de producir una creencia en la negación de P (1986: 111–112. *Cursivas nuestras*).

---

5 Cfr. al respecto Deweese–Boyd (2012).

El rasgo distintivo del intencionalismo, esto es, la presencia de una intención como condición necesaria para la producción de un estado de autoengaño, ha suscitado un extenso número de objeciones. El núcleo de estas objeciones puede reducirse a lo siguiente: la intención consciente de inducir en uno mismo una creencia falsa es un proyecto destinado al fracaso; el conocimiento de que planeamos creer aquello que consideramos falso resulta en un intento que se autoderrota. El autoengaño intencional, en síntesis, es imposible. Volvemos sobre estas objeciones en el apartado IV.

Pese al predominio de varias décadas de los modelos intencionalistas sobre el autoengaño, en los últimos años tales modelos han comenzado a ser desafiados por perspectivas que rechazan sus supuestos fundamentales. Estas perspectivas, denominadas a veces no intencionalistas (también “deflacionistas” o “motivacionistas”) suelen caracterizarse por la negación de los rasgos que distinguen a los enfoques clásicos del autoengaño (Barnes, 1997; Mele, 1997, 2001; Lazar, 1999; Van Leeuwen, 2007, 2008). Esto es, típicamente niegan que el autoengaño sea un fenómeno intencional y rechazan la tesis de la coexistencia de creencias contradictorias en el agente; en consecuencia, desaparece la necesidad de proponer una división de la mente como modo de lidiar con las paradojas resultantes del requisito anterior. El rechazo de tales compromisos obliga a proponer modelos explicativos que sugieran mecanismos alternativos a los propuestos por los intencionalistas.

Alfred Mele es sin duda el teórico más notorio entre los partidarios de los enfoques no intencionalistas. Caracteriza estos enfoques del siguiente modo:

La perspectiva anti-agencial: ninguna creencia motivacionalmente sesgada es producida o protegida intencionalmente. En ningún caso de creencia motivacionalmente sesgada de que *p* tratamos de causar la adquisición o retención de *p*, o tratamos de hacer más fácil para nosotros mismos adquirir o retener la creencia (2001: 13).

Las posiciones deflacionistas no requieren negar que puedan existir casos de autoengaño intencional; sólo requieren negar que en los casos típicos de autoengaño esté involucrada una intención como factor originador del proceso. Mele, en particular, sostiene que los ejemplos comunes de autoengaño, esto es, aquellos en los cuales alguien cree en algo que no debería creer sobre la base de las pruebas de que dispone, son reconocidos comúnmente como un fenómeno *motivado*: la persona cree aquello que le gustaría que fuese el caso. A este fenómeno puede dársele una interpretación no intencional, que es más sutil y menos problemática que aquella basada en el modelo proporcionado por el engaño interpersonal. Existen diversos

procesos que pueden contribuir en una variedad de modos con el sesgo motivado de las creencias, entre ellos la llamada “interpretación errónea negativa” (nuestro deseo de que  $p$  puede conducirnos a interpretar erróneamente datos que, en ausencia de tal deseo, consideraríamos adversos a  $p$ ; por el contrario, el deseo de que  $p$  nos lleva a considerar esos datos como favorables a  $p$ ) y la concentración/atención selectiva y recolección sesgada o selectiva de la evidencia. Asimismo, existen sesgos no motivados o “fríos” que actúan en la formación de creencias y que se combinan con los procesos recién mencionados, como la heurística de la disponibilidad y el sesgo de confirmación. Estos sesgos fríos o no motivados, si bien pueden funcionar independientemente de la motivación, pueden también ser provocados y sostenidos por ésta en la producción de creencias particulares motivacionalmente sesgadas. Por ejemplo, la motivación puede incrementar la vivacidad o saliencia de ciertos datos; los datos que avalan la verdad de una hipótesis que se desea que sea cierta pueden resultar más vívidos o salientes dado el reconocimiento de la dirección en la que apuntan. A lo anterior hay que agregar, con el fin de lograr una comprensión más completa del autoengaño, los modelos de testeo de hipótesis cotidianas, según los cuales la detección y minimización de errores cruciales es el principio organizador central en el testeo de tales hipótesis; las personas, según este modelo, son razonadores pragmáticos más interesados en minimizar errores cruciales que en llevar a cabo procesos de puesta a prueba tendientes a conocer la verdad. Quienes testean hipótesis comunes proceden como lo hacen en parte a causa de su deseo de evitar o minimizar “errores costosos”.

En cualquier caso, y como discutiremos más detalladamente en el apartado siguiente, para los deflacionistas el autoengaño no puede consistir típicamente en un proceso intencional y consciente, sino que requiere de medios indirectos que oculten al agente la naturaleza de su propio proyecto.

#### IV

Hemos señalado en el apartado inicial de este artículo que el esclarecimiento de la naturaleza de los vínculos entre el autoengaño y la inducción voluntaria indirecta de creencias requiere de un análisis más pormenorizado que el que Williams lleva a cabo. Como adelantamos, la existencia de perspectivas marcadamente diferentes sobre el autoengaño torna necesario un examen más minucioso acerca de las posibles relaciones entre ambos fenómenos.<sup>6</sup>

---

6 Uno de los árbitros de *Estudios de Filosofía* ha señalado la existencia de una ambigüedad en la tesis principal que defiende en este artículo. Esta ambigüedad, considera, reside en el hecho de que

De las distintas dimensiones del autoengaño que puedan tener relevancia para la discusión que nos ocupa, señalamos también, es sin duda central la referente a los procesos que conducen a la adquisición autoengañoso de una creencia. Dado que, como observamos en la sección II, la necesidad de este requisito ha generado uno de los mayores debates acerca del autoengaño, resulta conveniente profundizar en la caracterización de las posiciones descritas en el apartado precedente, el intencionalismo y el no intencionalismo. Comenzaremos por el primero.

Un examen de una de las estrategias que emplean los intencionalistas al intentar eludir algunos de los problemas generados por su concepción acerca del autoengaño permitirá, entiendo, comprender mejor las dificultades que plantea el requisito de intencionalidad.

Con el objetivo de ejemplificar la estrategia denominada “de la partición temporal”, ideada con el fin de defender al intencionalismo de la crítica relativa a que conduce a un estado mental imposible, Deweese–Boyd (2012) sugiere el siguiente ejemplo. Un oficial de policía involucrado en alguna conducta ilegal podría destruir cualquier rastro de esta conducta y crear pruebas que lo encubrirían (entradas en un diario, correos electrónicos y otras), sabiendo que probablemente olvidará el haber ejecutado esas conductas en unos pocos meses. Cuando, un año más tarde, sus actividades son investigadas, ha olvidado sus esfuerzos manipuladores y, basado

---

la afirmación de que el autoengaño no puede ser un medio eficiente para la inducción voluntaria indirecta de creencias no deja en claro si lo que se desea sostener es que el autoengaño no puede ser una condición necesaria, o necesaria y suficiente, o suficiente para tal inducción. Tal como intenté dejar en claro en el apartado I (y si mi interpretación es correcta), la tesis de Williams relativa a que el autoengaño es condición necesaria para la inducción voluntaria indirecta de creencias deja varias cuestiones inexploradas. Sin embargo, no ha sido mi objetivo refutar ninguna tesis específica de Williams. Tampoco ha sido mi objetivo refutar la pretensión (sea o no sostenida por Williams) de que el autoengaño es condición necesaria para la inducción voluntaria de creencias. Tal refutación, entiendo, sólo podría lograrse mediante un contraejemplo al modo intentado por Cook, es decir, un caso de tal inducción que, de modo inequívoco, no recurriera al autoengaño. Esto, entiendo, habida cuenta de los serios problemas (de los cuales he reseñado sólo unos pocos) que aquejan a cualquier caracterización del autoengaño, constituye una empresa cuya probabilidad de éxito es sumamente baja. Menos aún intento mostrar cómo el autoengaño no puede ser una condición necesaria y suficiente o sólo suficiente para la inducción voluntaria indirecta de creencias. Entiendo que el mostrar que el autoengaño no puede ser una condición necesaria para tal inducción puede revestir mayor interés filosófico que la tesis que aquí deseo defender, esto es, que el autoengaño no puede constituir un medio eficiente para la inducción voluntaria indirecta de creencias; sin embargo, habida cuenta de las dificultades que rodean a la noción de autoengaño, no creo que pueda ser posible. Agradezco al árbitro la observación de este punto, que me permitió corregir algunas afirmaciones en la versión previa de este artículo que sin duda contribuían a la ambigüedad por él señalada.



en tales pruebas fraguadas, cree falsamente que no ha estado involucrado en las actividades ilegales de las que es acusado. En este caso, observa Deweese–Boyd, el agente en cuestión no necesita sostener nunca de modo simultáneo creencias contradictorias, aun cuando ha intentado generar en sí mismo la creencia de que *p*, a la que considera falsa al inicio del proceso que conduce a su adopción, y verdadera al finalizar éste. Más aún, destaca, el agente ni siquiera necesita olvidar su intención original de generar el engaño; así, un no creyente que desea inducir en sí mismo la creencia en Dios bien podría recordar tal intención al finalizar el proceso y considerar que, por la gracia divina, incluso ese erróneo camino lo condujo a la verdad. En tales casos resulta crucial tener en cuenta que el éxito de la intención depende de la actuación de lo que se ha denominado “medios autónomos” (por ejemplo, la degradación normal de la memoria o la tendencia a creer lo que uno practica), no la continua conciencia de la intención.

Ahora bien, parecería que los casos anteriores no son lo que típicamente se consideraría casos de autoengaño intencional. Levy (2004) señala que no es éste el tipo de fenómeno que los defensores de la concepción tradicional tienen en mente. Si bien sostiene que el autoengaño es intencional, esta intencionalidad no es claramente perceptible para el agente; este no reconoce que está intentando engañarse. Para la concepción tradicional, quien se autoengaña se involucra en una conducta autoengañosa debido a que no desea enfrentar una verdad incómoda; pero admitir ante sí mismo que está tratando de engañarse implicaría el reconocimiento de la misma creencia que desearía evadir. En consecuencia, observa, “si el autoengaño es intencional, es una actividad intencional confusa y oscura” (p. 298). Pero el autoengaño que emplea medios indirectos como los descritos en el ejemplo precedente es claramente percibido; involucra la decisión deliberada de cambiar creencias que son conscientemente reconocidas. No es este el tipo de engaño hacia sí mismo, concluye Levy, aquel que los partidarios de la concepción tradicional considerarían autoengaño.

Si, como parecen sostener los intencionalistas, el autoengaño intencional es una actividad confusa y oscura, ¿cómo es posible ejecutar el proyecto de inducción de una creencia que se sabe falsa? Lazar (1999) observa que, más que explicar la formación de la creencia irracional, la presencia de una intención o plan para formar la creencia *constituye un impedimento hacia su formación*, dado que es probable que interfiera con el efecto de los mecanismos de sesgo. Por esta razón, las explicaciones de la formación de la creencia irracional que no apelen a una intención

de este tipo tienen *prima facie* una ventaja sobre la explicación intencionalista. El intencionalista puede no conceder este punto; podría afirmar que la intención de formar la creencia irracional no se interpone en el camino de la formación de la creencia en la medida en que el sujeto no sea consciente de ella. Lazar considera que esta réplica no puede salvar el enfoque intencionalista, dado que el intencionalista no puede responder de modo satisfactorio la pregunta de por qué el agente no es consciente de su intención de formar la creencia irracional.<sup>7</sup>

¿Tienen respuestas los intencionalistas ante las objeciones anteriores? Sin duda (Bermúdez, 2000; Nicholson, 2007). Pero más allá del éxito de estas réplicas subsiste un problema para su posición, que es el relativo a las probabilidades de éxito del intento de autoengaño; pospondremos el examen de esta dificultad y presentaremos primero algunas dificultades del no intencionalismo.

Hemos señalado que, para los no intencionalistas, el autoengaño es el resultado de la actuación de procesos cognitivos y emocionales/motivaciones que gozan de autonomía respecto de las intenciones del agente; la intención no es un componente necesario para generar un proceso conducente al autoengaño. Ahora bien, y más allá de tales premisas explicativas, resulta necesario preguntar cuál es para ellos el factor que inicia el proceso de adquisición autoengañoso de una creencia. Si no es una intención de adoptar la creencia preferida el factor que lo inicia, ¿cuál es este componente? Los defensores de esta posición no coinciden en la respuesta, por lo que hay varios candidatos: la ansiedad (Barnes, 1999), las emociones (Lazar, 1999), los deseos y los temores (Mele, 2001). Ahora bien, es posible preguntar entonces si toda vez que se encuentra presente, por ejemplo, un deseo intenso por un determinado estado de cosas en el mundo, estado que no tiene lugar, se producirá un proceso de inducción autoengañoso de la creencia de que tal estado ocurre u ocurrirá. La respuesta parece claramente negativa: si esto ocurriera, nos encontraríamos colectivamente en un estado de autoengaño generalizado. Parecería, entonces, que debe haber alguna condición adicional para que se inicie el proceso conducente a un estado de autoengaño.

La pregunta precedente plantea de modo tácito el denominado “problema de la selectividad”, que constituye posiblemente la crítica más difícil de responder por los deflacionistas. Esta objeción es presentada por Bermúdez (2000) en los siguientes términos. El autoengaño es paradigmáticamente selectivo; en consecuencia,

---

7 Cfr. Lazar (1999) para un examen detallado de las objeciones a la principal tesis intencionalista.

cualquier explicación de un caso determinado de autoengaño necesita explicar por qué los sesgos motivados actúan en *esa* situación particular. Sin embargo, el deseo de que  $p$  ocurra es insuficiente para activar los sesgos cognitivos en favor de la creencia de que  $p$ ; existen toda clase de situaciones en las cuales, independientemente de cuán fuerte sea nuestro deseo de que  $p$  ocurra, no experimentamos de ningún modo un sesgo en favor de la creencia de que  $p$ . La pregunta es entonces, concluye Bermúdez, cómo distinguir esas situaciones de otras en las cuales nuestros deseos resultan en un sesgo motivacional.

Tenemos, hasta aquí, que tanto intencionalistas como no intencionalistas tienen problemas a la hora de identificar y justificar los factores que dan origen al autoengaño: los primeros, por la dificultad de explicar la existencia de una intención de autoengaño de la cual el agente no puede ser consciente, y los segundos por el problema de determinar las condiciones suficientes que originen un estado de autoengaño. Sin embargo, lo anterior no responde a la pregunta relativa al grado en que estas dificultades se relacionan con la posibilidad del voluntarismo doxástico indirecto.

¿Se encuentran los intencionalistas en mejores condiciones que los deflacionistas para explicar la génesis del proceso que conduce a un estado de autoengaño? Sí, si le creemos a Bermúdez. La perspectiva intencionalista, señala, tiene una respuesta clara y directa al problema de la selectividad. La adquisición autoengañosa de que  $p$  requiere más que el simple deseo de que  $p$  ocurra y de un umbral bajo para la aceptación y alto para el rechazo de la hipótesis de que  $p$ . Requiere por parte de quien se autoengaña la intención de producir la adquisición de la creencia de que  $p$ . El hecho de que el intencionalista pueda resolver de este modo el problema de la selectividad parece, concluye Bermúdez, al menos una razón para pensar que no es posible abandonar enteramente una aproximación intencionalista acerca del autoengaño.

Independientemente del optimismo de Bermúdez acerca de las posibilidades de éxito de los enfoques intencionalistas acerca del autoengaño, parecería ser cierto que este fenómeno, bajo la perspectiva deflacionista, resultaría ser un medio menos indicado para la inducción voluntaria indirecta de creencias que entendido como una clase de proceso intencional. En efecto, si los modelos deflacionistas están en lo correcto, entonces el autoengaño típicamente no involucra intencionalidad. Ahora bien, si no hay intención de autoengaño, y no hay consecuentemente ningún intento por parte del agente de inducir en sí mismo ninguna creencia, parece muy difícil que

sea posible hablar de voluntarismo doxástico. Si la creencia que se desea adquirir intencionalmente, pero por un medio indirecto como el autoengaño, está sujeta a procesos que no se encuentran casi en absoluto bajo nuestro control volitivo, el voluntarismo doxástico indirecto parece simplemente imposible.

Lo anterior se encuentra reforzado por nuestro conocimiento de los procesos de distorsión a los que se encuentra sujeta la adopción de creencias. Thagard (2011) identifica varias decenas de sesgos que afectan tanto nuestras creencias como nuestras acciones, entre ellos varios bien conocidos, como la falacia de la conjunción, el sesgo de confirmación, el exceso de confianza en el propio juicio y la tendencia a ignorar la retroalimentación proveniente de la experiencia. Se ha discutido (Levy, 2004) la medida en la cual podemos ser conscientes de estos sesgos y, hasta cierto punto, controlar su acción. Sin embargo, no es posible, hasta donde sabemos, actuar voluntariamente de modo tal que estos sesgos se activen en una dirección determinada y, menos aun, generar a través de un grupo definido de ellos un estado de autoengaño.

Lo dicho vale no sólo para el deflacionismo, como hemos indicado, sino también para el intencionalismo: la intención de sesgar nuestra elección de creencias no puede afectar de un modo completamente direccionado, por su propia naturaleza, a sesgos implícitos que se encuentran claramente más allá del control volitivo.<sup>8</sup> Sin embargo, cabe preguntar si el autoengaño, entendido a la manera en que postulan los intencionalistas, ofrece realmente mejores probabilidades para la inducción voluntaria de creencias. Aun cuando admitamos que algunos casos de autoengaño puedan ser generados intencionalmente, no es obvio que la combinación de intencionalidad y procesos sesgados conduzca invariablemente a la adquisición de la creencia deseada; lo contrario supondría el control de una compleja serie de procesos implícitos y automáticos que se encuentran fuera de nuestro alcance volitivo. Pero si esto fuese así, entonces el rol del autoengaño en la sustentación del voluntarismo doxástico indirecto debería ser puesto en tela de juicio; esto no implica sostener que este fenómeno es imposible, sino que es como mínimo muy dudoso que el autoengaño pueda constituir un medio eficiente para la adquisición voluntaria indirecta de creencias.

---

8 Davidson (1986) reconoce que hay un “buen número” de mecanismos que tienen lugar en el autoengaño (por ejemplo, “desplazar hacia el fondo la evidencia negativa o a destacar la positiva”: 113); sin embargo, su interés no es el de identificar tales mecanismos, sino argumentar en favor de la necesidad de que el agente sea consciente de que la evidencia apoya la creencia contraria a sus deseos.

Como vimos al inicio de este trabajo, es innegable que poseemos cierta capacidad para influir voluntariamente en nuestras creencias mediante medios más o menos indirectos. Estos medios incluyen, por ejemplo, la estrategia de exponernos a ciertas fuentes de información y no a otras, lo que refuerza algunas de nuestras creencias por medio de la ausencia de crítica o información discrepante. Ahora bien, el control voluntario indirecto sobre las creencias parece algo distinto, ya que parece requerir a su vez cierto grado de control sobre los medios que ponemos en juego para el logro de ese objetivo. Si las precedentes observaciones sobre los procesos que originan y producen el autoengaño son correctas, este fenómeno no puede constituir un medio adecuado ni menos aun eficiente para inducir en nosotros las creencias que buscamos. ¿Controlamos indirectamente nuestras creencias si los medios que ponemos en juego no garantizan el objetivo, sino que en el mejor de los casos incrementan en una medida no determinada la probabilidad de lograrlo? Diría que la respuesta es negativa; más que un control indirecto, parecería tratarse de una apuesta con mayor o menor grado de riesgo.

### Bibliografía

1. Adler, J. (2002). *Belief's Own Ethics*. Cambridge, The M.I.T. Press.
2. Bach, K. (1981). An Analysis of Self Deception. *Philosophy and Phenomenological Research*, 41, (3), 351–370.
3. Barnes, A. (1997). *Seeing through self-deception*. New York: Cambridge University Press.
4. Bennett, J. (1990). Why Is Belief Involuntary? *Analysis*, 50, (2), 87–107.
5. Bermúdez, J. L. (2000). Self Deception, intentions and contradictory beliefs. *Analysis* 60, (4), 309–319.
6. Booth, A. R. (2007). Doxastic voluntarism and self-deception. *Disputatio*, II, 2 (22), 115–130.
7. Cook, J. T. (1987). Deciding to Believe Without Self-Deception. *The Journal of Philosophy*, 84, (8), 441–446.
8. Davidson, D. (1982). Paradoxes of irrationality. En Richard Wollheim & James Hopkins (eds.). *Philosophical Essays on Freud* (pp. 289–305) New York: Cambridge University Press.

9. Davidson, D. (1986). Engaño y división. En *Mente, mundo y acción*. Barcelona: Paidós.
10. Demos, R. (1960). Lying to Oneself. *Journal of Philosophy*, 57, 588–95.
11. Deweese–Boyd, I. (2012). Self Deception. Stanford Encyclopedia of Philosophy. Extraído el 10/06/13 de. <http://plato.stanford.edu/entries/self-deception/>
12. Elster, J. (1979). *Ulysses and the Sirens*. New York: Cambridge University Press.
13. Funkhouser, E. (2003). Willing Belief and the Norm of Truth. *Philosophical Studies*, 115, 179–195.
14. Funkhouser, E. (2005). Do the Self Deceived Get what they Want? *Pacific Philosophical Quarterly*, 86, 295–312.
15. Gardiner. (1970). Error, Faith and Self Deception. *Proceedings of the Aristotelian Society*, 70 (1), 221–224.
16. Lazar, A. (1999). Deceiving Oneself or Self–Deceived? On the Formation of Beliefs ‘Under the Influence’. *Mind* 108, (430), 265–290.
17. Levy, N. (2004). Self Deception and Moral Responsibility. *Ratio (new series)*, XVII., 294–311.
18. Lynch, K. (2010). Self–Deception, Religious Belief, and the False Belief Condition. *The Heythrop Journal*, LI., 1073–1074.
19. Mele, A. (1987). *Irrationality. An Essay on Akrasia, Self–Deception, and Self–Control*. New York–Oxford: Oxford University Press.
20. Mele, A. (1997). Real Self Deception. *Behavioral and Brain Sciences* 20, (1), 91–136.
21. Mele, A. (2001). *Self–deception Unmasked*. Princeton, Princeton: University Press.
22. Nicholson, A. (2007). Cognitive Bias, Intentionality and Self–Deception. *Teorema* XXVI (3). 45–58.
23. Nottelman, N. (2006). The Analogy Argument for Doxastic Voluntarism. *Philosophical Studies*, 131, 559–582.

24. Oksenberg–Rorty, A. (1988). The Deceptive Self: Liars, Layers and Lairs. En: Brian McLaughlin & Amelie Oksenberg–Rorty (eds.). *Perspectives on Self–Deception*. University of California Press.
25. Paluch, S. (1967). Self–Deception. *Inquiry*, 10, 268–78.
26. Pears, D. (1985). The goals and strategies of self–deception. En Jon Elster (ed.), *The Multiple Self*. New York: Cambridge University Press.
27. Ryan, S. (2003). Doxastic compatibilism and the ethics of belief. *Philosophical Studies*, 114, 47–79.
28. Santamaría Navarro, V. (2009). Beliefs: The Will Besieged by the Evidence. *Teorema*, XXIII (3), 131–149.
29. Sartre, J. P. (1943). *El Ser y la Nada*. Buenos Aires: Losada.
30. Steup, M. (2008). Doxastic freedom, *Synthese*, 161, 375–392.
31. Szabados, B. (1974). Self–Deception. *Canadian Journal of Philosophy*, 4, 51–68.
32. Talbott, W. (1995). Intentional Self–Deception in a Single Coherent Self. *Philosophy and Phenomenological Research* 55, (1), 27–74.
33. Thagard: (2011). Critical Thinking and Informal Logic: Neuropsychological Perspectives. *Informal Logic* 31, (3), 152–170.
34. Van Leeuwen, D. S. N. (2007). The Product of Self–Deception. *Erkenntnis*, 67, 419–437.
35. Van Leeuwen, D. S. N. (2008). Finite rational self–deceivers. *Philosophical Studies*, 139, 91–208.
36. Van Leeuwen, D. S. N. (2009). Self–Deception Won’t Make You Happy. *Social Theory and Practice*, 35, (1), 107–132.
37. Williams, B. (1973). Deciding to believe. En *Problems of the Self*. *Philosophical Papers 1956–1972*. Cambridge: Cambridge University Press.