

ARTICULOS/ARTICLES

**SEIS DÉCADAS DE ESTUDIOS SOBRE EL AUTOENGAÑO:
PROBLEMAS PERENNES Y NUEVOS INTERROGANTES***

**SIX DECADES OF STUDIES ON SELF-DECEPTION: PERENNIAL
PROBLEMS AND NEW QUESTIONS**

Gustavo Fernández Acevedo
Facultad de Humanidades
Universidad Nacional de Mar del Plata

Resumen:

En las últimas seis décadas el fenómeno del autoengaño ha sido objeto de creciente interés no sólo en el ámbito de la filosofía, dentro de la cual fue tradicionalmente estudiado, sino también en el de distintas ciencias, entre ellas la psicología, las neurociencias, la biología evolucionista y las ciencias sociales. Este incremento en el interés ha redundado en una proliferación de interrogantes y propuestas teóricas de muy diversas clases, sin que hasta la fecha se haya logrado una teoría unificada que goce de consenso respecto del fenómeno. En el presente artículo se presentan las preguntas que contemporáneamente se consideran más relevantes sobre el autoengaño, tanto dentro del dominio de la filosofía como en el de diversas ciencias. Asimismo, se analizan algunos problemas conceptuales y epistemológicos derivados de la proliferación de estudios sobre el fenómeno, y se examinan algunas de las líneas de investigación más recientes y promisorias desarrolladas con el fin de resolver algunas dificultades conceptuales bien conocidas.

Palabras clave: Autoengaño; Filosofía; Ciencia; Explicación; Creencia.

Abstract:

In the last six decades the phenomenon of self-deception has been subject of increasing interest not only in the philosophical domain, in which classically

* Deseo agradecer las observaciones de los árbitros de *Páginas de Filosofía*, que me permitieron mejorar algunos aspectos de este artículo. Identifico y respondo tales observaciones en notas al pie.

was studied, but in several sciences too, including psychology, neurosciences, evolutionary biology and social sciences. This increased interest has resulted in a plenty of questions and theoretical proposals, but we are at the present far from a unified theory that enjoys of general acceptance. In this article I introduce the most contemporarily relevant questions about self-deception, both in philosophy as in science. In addition, I analyze several conceptual and epistemological problems raised by the proliferation of theories about self-deception, and examine some more recent and promising research lines that intend to solve some well-known conceptual difficulties.

Key Words: Self-deception; Philosophy; Science; Explanation; Belief.

I

Transcurridas casi seis décadas desde la publicación del artículo pionero de Raphael Demos “Lying to Oneself” (1960), el interés por el fenómeno del autoengaño, lejos de extinguirse, se ha mantenido o incluso incrementado. Cientos de artículos en revistas especializadas y un número considerable de libros sobre el problema atestiguan esta vigencia. Este sostenido interés, sin embargo, no se ha visto coronado por el desarrollo de una teoría unificada sobre el fenómeno que goce de consenso entre los especialistas, y menos aún de aceptación unánime; más aún, podría decirse que prácticamente ningún aspecto del problema del autoengaño ha perdido el carácter de controversia persistente. De hecho, la propia caracterización del fenómeno es objeto de debate. A los fines expositivos, propondremos algunos rasgos que parecen caracterizar al autoengaño; tal caracterización puede ser presentada como la conjunción de tres condiciones, a saber:

- a. adquisición y/o mantenimiento de una creencia falsa,
- b. frente a elementos de juicio contrarios a tal creencia,
- c. motivados por procesos mentales no cognitivos (por ejemplo, deseos o emociones) que favorecen la adquisición y/o retención de esa creencia.¹

¹ Si bien todos los componentes de esta caracterización pueden ser cuestionados por distintas razones, ya sea porque presupone demasiado (que resulte necesario que la creencia sea falsa), ya sea porque resulta insuficiente (según algunos autores, el autoengaño es un fenómeno típicamente intencional, rasgo que no aparece en ella), el desarrollo de estos posibles cuestionamientos excedería los alcances de este trabajo.

La expansión de los estudios sobre el problema presenta ramificaciones de distinto tipo, que hacen más difícil el reconocimiento del territorio conceptual. Una primera ramificación está constituida por la captación del problema por parte de disciplinas que hasta el momento habían tenido una importancia nula o marginal en lo que respecta a sus aportes teóricos. La psicología, la biología evolucionista, la neurología y la neuropsicología y las ciencias sociales comenzaron a ocuparse de modo sistemático de su estudio. De este modo, la filosofía cedió su virtual monopolio sobre el problema del autoengaño, que se convirtió así en un territorio explorado por distintas disciplinas, exploración por lo general no libre de conflictos ni menos aún coordinada. Una segunda ramificación está dada por el planteo de interrogantes nuevos, que se sumaron a los más tradicionales y fundamentales, esto es, las preguntas relativas a la naturaleza del autoengaño, sus condiciones de posibilidad y las razones de su existencia, así como su relación con otras formas de irracionalidad motivada, como la *akrasia* y el pensamiento desiderativo. Tales interrogantes nuevos incluyen, entre otros, los siguientes: la posibilidad de formas de autoengaño colectivo; la relación entre autoengaño y creencia religiosa; los vínculos existentes entre el autoengaño y fenómenos psicopatológicos, como el delirio y la confabulación; las consecuencias del autoengaño para la salud psíquica y la felicidad; y sus posibles ventajas adaptativas desde una perspectiva evolucionista.

Esta proliferación tanto disciplinar como teórica obliga, si se pretende establecer algún orden conceptual, a sugerir un criterio organizador. Este criterio organizador puede estar constituido, entiendo, por una serie de preguntas que admiten casi invariablemente una respuesta dicotómica y que apuntan a aspectos nucleares del autoengaño, tanto en lo que respecta a su explicación teórica como, al menos en algunos casos, a sus consecuencias prácticas. Esta serie de preguntas incluye las que, a mi modo de ver, son las preguntas a las que se les ha dado más importancia en las últimas seis décadas. Todas ellas (con la posible excepción de la pregunta por la responsabilidad moral por el autoengaño, que ha sido de manera más exclusiva de interés filosófico) han sido objeto en mayor o menor medida tanto de estudios filosóficos como empíricos.

Ahora bien, como se señaló, la proliferación de teorías e interrogantes novedosos sobre el autoengaño no ha implicado un avance definido hacia una teoría unificada o hacia cierto consenso respecto de algunas posibles respuestas; por el contrario, han surgido problemas

conceptuales nuevos, cuyo análisis constituye la otra parte de los objetivos de este artículo. De los posibles problemas conceptuales y epistemológicos planteados por la proliferación de estudios sobre el autoengaño nos referiremos a dos en particular: el primero, relativo a la pérdida de precisión conceptual observable en diversos estudios sobre el problema; el segundo, derivado de la coexistencia de múltiples explicaciones del fenómeno, provenientes de distintas disciplinas científicas.

El artículo está estructurado de la siguiente forma. En la segunda sección expondré cada una de las preguntas fundamentales sobre el autoengaño, junto con las principales respuestas que se han ofrecido; cuando resulte pertinente señalaré, además, las relaciones conceptuales o empíricas entre las distintas respuestas. En la tercera sección presentaré lo que, a mi modo de ver, constituyen déficits en algunos análisis contemporáneos del problema, derivados del descuido de precisiones conceptuales provenientes habitualmente de la reflexión filosófica. En la cuarta sección describiré un problema epistemológico derivado de la proliferación de respuestas a los distintos interrogantes relativos al autoengaño, basándome para ello en la existencia de distintos tipos de propuestas explicativas del autoengaño. Por último, en la quinta sección describiré algunas posibles líneas de investigación potencialmente promisorias para los estudios venideros sobre el problema.

II

1. ¿Es posible el autoengaño?

Una de las primeras afirmaciones que hemos hecho es que virtualmente cualquier aspecto del autoengaño, incluyendo su propia existencia, ha sido objeto de debate. En consecuencia, tenemos aquí tenemos una primera división de autores según sea su posición respecto de la existencia del fenómeno. En este caso, y como excepción a la naturaleza dicotómica a la que se hizo referencia en relación a las restantes preguntas fundamentales, creo que es útil distinguir tres posiciones, y no sólo dos como en ocasiones se hace.

En primer lugar, tenemos que las posiciones que podemos llamar “escépticas” respecto del autoengaño. Diversos estudiosos han negado la existencia de este fenómeno (Paluch, 1967; Bok, 1980; Kipp, 1980; Bandura; 1991). Si bien esta posición ante el problema del autoengaño parecería haber perdido impulso en los últimos años, su vigencia no ha desaparecido por completo (Borge, 2003). Este escepticismo suele

sustentarse en la aparente existencia de paradojas presuntamente insolubles de distintas clases. Una de las dos más estudiadas es la llamada en ocasiones “paradoja doxástica” del autoengaño. Sintéticamente: ¿cómo es posible que alguien crea simultáneamente p y $\neg p$? En opinión de los escépticos, ninguno de los intentos para lidiar con este problema ha tenido ni puede tener éxito, por lo que sólo cabe concluir que el autoengaño no existe.

Una segunda posición es la que podría denominarse, “disolucionista” respecto del autoengaño (Harré, 1988; Lewis, 1996; Clegg y Moissinac, 2005). Esta perspectiva se inspira en las corrientes de pensamiento posmodernas y construccionistas sociales, y reconoce de manera inequívoca la influencia de K. Gergen. No se trata de una posición análoga a la anterior, que sostiene que el autoengaño es imposible; lo que se afirma es que el autoengaño es simplemente una cuestión de perspectivas diferentes, algo que está más en la mente del observador que un proceso intrapsíquico en el presunto autoengañado, una cuestión cultural más que un proceso psicológico de alguna clase.

Por último, tenemos una abrumadora mayoría de autores que aceptan la existencia del autoengaño como un problema legítimo a resolver: Demos (1960); Fingarette (1969); Bach (1981); Audi (1982); Davidson (1985); Mele (1987, 1997, 2001, 2003); Johnston (1988); McLaughlin (1988); Oksenberg-Rorty (1988); Barnes (1997); Talbot (1995); Lazar (1999); Scott-Kakures (2000, 2001, 2009); Trivers (2000, 2010); Funkhouser (2005); Van Leeuwen (2007a, 2007b, 2008, 2009), y muchos otros. Para estos autores no se trata de un seudoproblema originado o bien en un uso descuidado del lenguaje, que nos induce al error de suponer la existencia de un fenómeno donde sólo hay un término vacío o mal utilizado (como sostienen los escépticos), o bien en una concepción tradicional y errónea acerca de la mente (como afirman los disolucionistas). Sólo una vez que se admite la existencia del autoengaño es posible plantear los interrogantes que presentaremos a continuación.

2. ¿Es intencional?

La posición clásicamente defendida por los filósofos concibe al autoengaño como un fenómeno intencional (Davidson, 1985; Oksenberg-Rorty, 1988; Pears, 1991; Talbot, 1995; Bermúdez, 2000). Según los modelos intencionalistas, lo que ocurre cuando alguien se engaña a sí mismo es estructuralmente similar a lo que ocurre cuando una persona engaña intencionalmente a otra, con la obvia diferencia de que en el primer caso la persona se engaña *intencionalmente* a sí misma. Cabe

aclarar que los modelos intencionalistas no suponen una intención clara y consciente por parte del agente de cambiar una creencia verdadera por una falsa; esto último supondría la aceptación del voluntarismo doxástico directo, posibilidad habitualmente reconocida como imposible (Williams, 1973; Adler, 2002).² El componente intencional del autoengaño, en caso de existir, debe ser indirecto u oblicuo. El enfoque intencionalista, en consecuencia, concibe al sujeto autoengañado como alguien que actúa con la intención de que esa acción cause la formación de la creencia deseada; tal acción puede consistir en un direccionamiento intencional de la atención lejos de las pruebas que apoyan a *p* (la creencia que rechaza), o puede consistir en una búsqueda activa de elementos de juicio contra *p*. En segundo lugar están aquellos que sostienen que es una clase de fenómeno no intencional (Mele, 1987, 2001; Barnes, 1997; Lazar, 1999; Levy, 2004). De acuerdo con las concepciones no intencionalistas, la formación de creencias autoengañosas puede ser explicada sin recurrir a la acción intencional. El autoengaño se explicaría por una combinación de sesgos cognitivos y factores motivacionales y afectivos; no haría falta postular una intención de autoengañarse por parte del agente.

3. ¿Implica la coexistencia de creencias contradictorias?

Este problema está íntimamente relacionado con las concepciones intencionalistas que, como se observó, han constituido durante muchos años la posición estándar respecto del autoengaño. Si el autoengaño puede comprenderse a partir del modelo proporcionado por el engaño interpersonal, entonces parecería que debemos admitir que, como ocurre en aquél, el agente que se autoengaña posee tanto la creencia verdadera, que rechaza, como la creencia falsa, que es la que adquiere y/o mantiene (Demos, 1960; Davidson, 1985; Gur y Sackheim, 1979). Esto, como es dable imaginar, conduciría a una situación cognitiva imposible: la coexistencia en la conciencia de una contradicción visible llevaría inevitablemente a la eliminación de una de las creencias. Una de las alternativas de solución típicas ha sido la de postular una división o partición de la mente que evite que las creencias incompatibles entren en contacto. La otra es negar que las creencias contradictorias deban coexistir (Mele, 2001; Barnes, 1997; Talbot, 1995). Esto es, para que exista el autoengaño el agente sólo debe poseer la creencia falsa, en

² Aunque véanse Bennet (1990) Ryan (2003) y Steup (2008) para réplicas a las objeciones al voluntarismo doxástico directo y defensa de esta posición.

presencia de pruebas favorables a la creencia verdadera. Se han considerado al menos dos alternativas para defender esta perspectiva: o bien el agente sostuvo inicialmente y luego desechó la creencia verdadera en favor de la creencia falsa (con lo cual las creencias contradictorias no llegarían a coexistir), o bien no hace falta que el agente haya desestimado, y ni siquiera poseído inicialmente, la creencia verdadera, sino que adopta inicialmente la creencia falsa. Sean o no admisibles estas alternativas, la posible existencia de creencias contradictorias ha constituido un tema de debate recurrente en torno del autoengaño.

4. ¿Su explicación requiere de una división de la mente?

Este debate tiene lugar entre quienes afirman que la existencia del autoengaño requiere de una división del psiquismo (Demos, 1960; Davidson, 1985; Oksenberg-Rorty; 1988; Pears, 1991), y aquellos que sostienen que tal división no es necesaria (Mele, 2001; Barnes, 1997; Talbot, 1995). Como mencionamos, las posiciones divisionistas suelen recurrir a la tesis de una partición o división de la mente como forma de solucionar el problema generado por la presunta coexistencia de creencias contradictorias. Esta división de la mente puede coincidir con alguna teoría empírica acerca del funcionamiento mental (por ejemplo, alguna teoría cognitiva o neuropsicológica acerca de procesos conscientes e inconscientes), o bien constituir una teoría *ad hoc*. Un punto importante a destacar aquí es que muchas teorías divisionistas, como es dable imaginar, son cuestionadas por introducir distinciones que, en el mejor de los casos, solucionan un problema al costo de generar otro de igual o mayor dificultad. Por el contrario, las teorías no divisionistas parecen tener a su favor el ser más parsimoniosas, al no proponer ninguna división adicional de la mente a aquellas requeridas para explicar otros fenómenos mentales distintos del autoengaño.

5. ¿Es adaptativo?

Lo primero que es necesario hacer respecto de esta pregunta es aclarar en qué sentido se entiende el término “adaptativo”. A veces, dentro del campo de la Psicología, se emplea el término “adaptación” en un sentido relacionado con el ajuste del individuo a su medio social.³

³ Véase, por ejemplo, el empleo que se hace de “adaptativo” en el Manual Diagnóstico y Estadístico de los Trastornos Mentales, Cuarta Edición, en referencia al denominado “trastorno adaptativo”: “La característica esencial del trastorno adaptativo es el desarrollo de síntomas emocionales o comportamentales en respuesta a un estresante

También, como han hecho algunos psicólogos, puede emplearse para hacer referencia a alguna clase de proceso útil para la vida y productor de salud mental. Muchos autores, antes de que se comenzara a estudiar de manera científica el autoengaño, señalaron su carácter protector frente a las verdades dolorosas y a los sufrimientos vitales (Williams, 1973; Davidson, 1985). Más allá de este uso posible del término “adaptativo”, el significado del término al que queremos hacer referencia ahora es el proporcionado por la teoría de la evolución. “Adaptativo” hace referencia a un rasgo de un organismo que contribuye con su aptitud, esto es, con su capacidad para producir descendencia, perpetuando así su patrimonio genético. La pregunta, entonces, es si la capacidad para el autoengaño constituye un rasgo seleccionado evolutivamente por su contribución con nuestra aptitud (Trivers, 2000; Von Hippel y Trivers, 2011; Surbey, 2011), o si se trata de alguna clase de subproducto estructural (Van Leeuwen, 2007b, 2008; Kurzban-Aktipis, 2007), esto es, de un rasgo no seleccionado pero asociado con otros rasgos que sí han sido seleccionados por tal contribución.

6. ¿Contribuye con el bienestar psíquico y/o felicidad?

Esta es quizás la pregunta más moderna y en apariencia paradójica de las siete que hemos seleccionado. Si bien, como se mencionó en el apartado precedente, el autoengaño parece tener una función de protección contra verdades dolorosas, tradicionalmente se ha considerado que es una clase de error censurable, tanto desde el punto de vista de la racionalidad como desde la perspectiva ética. Claramente, atenta contra la máxima tradicional del Oráculo de Delfos: “conócete a ti mismo”. Sin embargo, hace ya alrededor de tres décadas comenzaron a aparecer una serie de trabajos que cuestionaron la visión tradicionalmente negativa sobre el autoengaño. En particular, los trabajos de S. Taylor y sus colaboradores enfatizaron la importancia de lo que denominaron “ilusiones positivas” para hacer referencia a una serie de creencias que no son sostenidas por los elementos de juicio (autoevaluaciones positivas no realistas, percepciones exageradas de control o habilidad y optimismo no realista) y que pueden servir a una amplia variedad de funciones

psicosocial identificable. Los síntomas deben presentarse durante los 3 meses siguientes al inicio del estresante (Criterio A). La expresión clínica de la reacción consiste en un acusado malestar, superior al esperable dada la naturaleza del estresante, o en un deterioro significativo de *la actividad social o profesional (o académica)* (Criterio B)” (DSM IV, p. 639, cursivas nuestras).

cognitivas, afectivas y sociales. Plantean también el intento de resolver una paradoja: ¿cómo pueden las percepciones erróneas de uno mismo y del entorno ser adaptativas cuando procesar la información con precisión parece ser esencial para el aprendizaje y el funcionamiento exitoso en el mundo? Este punto de vista (Taylor y Brown, 1988; Trivers, 2000; Erez et al, 1995) ha sido tan influyente como polémico; otros autores (Colvin et al, 1995; Badhwar, 2008; Van Leeuwen, 2009; Martin, 2012) han objetado que los presuntos bienestar y felicidad obtenidos por medio del autoengaño, en el dudoso caso de que fuesen posibles, serían ilusorios o efímeros.

7. ¿Implica responsabilidad moral?

Como es fácil de imaginar, esta pregunta ha sido una de las de mayor interés dentro del ámbito filosófico, y sin duda es aquella a la que se le ha dado la respuesta que goza del mayor consenso entre los estudiosos del tema. Este consenso, prácticamente unanimidad, ha consistido en la afirmación de que quien se autoengaña es responsable y censurable por su estado; es digno de reproche desde el punto de vista moral (Butler, 1726; Baron, 1988; Mele, 2001; Deweese-Boyd, 2008). Es de imaginar, también, que existe una conexión muy directa entre las concepciones intencionalistas del autoengaño y la posibilidad de atribuir responsabilidad moral: si alguien se autoengaña intencionalmente (aun cuando sea de manera oblicua e indirecta), entonces tiene control sobre sus acciones y, en consecuencia, es responsable por ellas. Sin embargo, esta situación ha comenzado a no ser tan clara a partir de los enfoques “deflacionistas” del autoengaño. Si el autoengaño no es producto de alguna clase de intención del agente, sino de una compleja combinación de sesgos cognitivos no conscientes y estados emocionales y motivacionales, parece mucho más difícil, a excepción de ciertas circunstancias muy específicas, adjudicarle responsabilidad moral (Levy, 2004).

El territorio conceptual configurado principalmente por estas siete preguntas y las respuestas que a ellas se han ofrecido permite, a nuestro entender, comprender tanto el panorama de los estudios contemporáneos sobre el autoengaño como los derroteros teóricos que su examen ha adquirido en las últimas décadas. Sin embargo, esta ganancia en amplitud e interés se ha visto en alguna medida contrapesada, como mostraremos en las dos próximas secciones, por deficiencias conceptuales de algunas

de las respuestas, por una parte, y por una sobreabundancia explicativa, por la otra.⁴

III

Como adelantamos en la primera sección, la posibilidad de establecer un conjunto de interrogantes que sinteticen los principales intereses tanto filosóficos como científicos acerca del autoengaño no equivale al logro de un consenso sobre las posibles respuestas. Más aún, el avance en el planteo de problemas científicos sobre el fenómeno ha estado acompañado, en ocasiones, por el descuido de precisiones conceptuales que incluso llegan a generar dudas acerca de la medida en que tales estudios constituyen aportes genuinos para nuestra comprensión del autoengaño. Existen diversos ejemplos de este proceso.

Paulhus (2007), desde la psicología social, define al autoengaño como “el acto de mentirse a uno mismo” (p. 801). Refiere a esta conducta como casos en los cuales las “personas aparentemente creen algo que saben que es falso” (p. 801). Aclara que tal conducta no incluye

⁴ Los árbitros de *Páginas de Filosofía* han objetado que el artículo no siempre logra aunar los argumentos filosóficos con su base empírica, ya que no se muestra con claridad cuál es ésta en los distintos casos, así como la disciplina de la que proviene. Entiendo que la objeción tal vez se deba a que la formulación empleada en el texto puede dar lugar a la impresión de que existe una base empírica para cada uno de los diferentes argumentos filosóficos, como observan los árbitros. Cabe aclarar al respecto que no supongo (aunque no presumo poseer la totalidad de la evidencia pertinente) que todos los argumentos filosóficos sintetizados en el texto se sustenten en una base empírica proveniente de alguna ciencia fáctica. Esto se pone de manifiesto nítidamente, entiendo, en los argumentos escépticos acerca del autoengaño: la imposibilidad de este fenómeno, según tales argumentos, es conceptual, no fáctica: no hay una manera conceptualmente plausible de dar sentido a la noción intuitiva de que alguien se engaña a sí mismo. Si, ahora bien, se admite la existencia del fenómeno, algunos de los argumentos relativos a él (como su carácter intencional, el requisito de la coexistencia de creencias contradictorias y, en menor medida, la tesis de la división de la mente), de hecho, parecen provenir de exigencias conceptuales debidas a la modelización del autoengaño sobre la base provista por el engaño interpersonal, y el recurrir a conceptos provenientes de determinadas teorías empíricas –por ejemplo, el psicoanálisis– es una opción no forzosa desde el punto de vista teórico. La situación es distinta, no obstante, con algunas posiciones filosóficas sobre el autoengaño que hacen uso en mayor (por ejemplo, en el caso de Mele) o menor medida de información empírica aportada por la psicología cognitiva. Los debates relativos al carácter adaptativo del autoengaño y sus potenciales beneficios por su parte, se originan en cuestiones empíricas, aunque claramente, en el caso de sus eventuales beneficios, no es posible dejar de lado la dimensión filosófica.

exageración, falsificación o simple mentira; el autoengaño es algo más profundo y complicado, incluso paradójico. Ejemplos de este fenómeno son los casos de un hombre habitualmente agradable que bebe demasiado y rechaza la creencia de que tiene un problema con el alcohol, o la madre de un criminal que rechaza lo que la policía dice de él. Lo que caracteriza a todos estos casos es que la evidencia en favor de la creencia rechazada es mayor que la evidencia contraria a ella. Paulhus considera que la explicación de casos como éstos requiere del reconocimiento de la existencia de partes inconscientes de la mente. Sólo en tales partes inconscientes puede un conflicto emocional influenciar realmente la conducta de un individuo y, pese a eso, ser inaccesible.

Así como a Paulhus la coexistencia de creencias contradictorias y la postulación de una concepción agencial del fenómeno (esto es, las personas somos agentes activos en la producción del autoengaño) le parecen perfectamente aceptables, para otros psicólogos estas mismas tesis son tan evidentemente incorrectas como para rechazar la posibilidad del autoengaño. Bandura (1991, 2011) ha negado de manera tajante la existencia de este fenómeno. Basa su posición escéptica en la conocida tesis de la imposibilidad de ser simultáneamente quien engaña y quien es engañado; es lógicamente imposible, señala, engañarse a uno mismo para creer algo mientras simultáneamente se sabe que es falso. Observa, asimismo, que los intentos de resolver esta paradoja mediante la apelación a partes inconscientes de la mente han tenido escaso éxito. Considera que esta estrategia, más que explicar el autoengaño, lo aniquila. Estas concepciones del yo dividido, como las denomina, fracasan al tener que explicar como un yo conciente puede mentir a un yo inconciente; el yo engañador tiene que ser conciente de lo que el yo engañado cree para saber como fraguar el engaño.

Un tercer ejemplo de adopción de definiciones cuestionables se encuentra en los intentos experimentales (por otra parte muy destacados) de probar la existencia del autoengaño. Lugar de privilegio en esta vertiente es ocupado por los estudios de Gur y Sackheim (1979). Estos autores proponen los siguientes criterios como necesarios y suficientes para el autoengaño: “1. el individuo sostiene dos creencias contradictorias (que *p* y que no *p*). 2. Esas dos creencias contradictorias son sostenidas simultáneamente. 3. El individuo no es conciente de que sostiene una de esas creencias. 4. El acto que determina cuál creencia es objeto de la conciencia y cuál no es un acto motivado” (p. 149).

Sin embargo, no sólo en estudios propiamente psicológicos pueden hallarse caracterizaciones cuestionables del fenómeno. Un caso

de descuido de mínimas precisiones conceptuales puede encontrarse en Sharma et al (2010). Estos autores señalan que el engaño implica un esfuerzo consciente del sistema nervioso central, acompañado por reacciones fisiológicas del sistema autónomo que no se encuentran bajo el control consciente, y caracterizan el autoengaño, citando a Kant, como “la enfermedad de la falsedad”. Un rasgo esencial del autoengaño es, sostienen, que dos creencias contradictorias son sostenidas simultáneamente pero el agente no es consciente de su existencia. El autoengaño, concluyen siguiendo a Gur y Sackheim, constituye un acto inconsciente.

Los estudios reseñados son susceptibles de una misma clase de objeción: el partir de caracterizaciones como mínimo discutibles del fenómeno del autoengaño. Esto se debe a que, como vimos en la segunda sección, todos comparten un mismo supuesto cuestionable: el de que el autoengaño debe implicar esencialmente la coexistencia de creencias contradictorias. La adopción de este supuesto conduce, a su vez, a dos alternativas igualmente innecesarias: o bien, como es el caso de Paulhus, Gur y Sackheim y Sharma et al, a proponer partes inconcientes de la mente de una forma *ad hoc*, o bien, como es el caso de Bandura, a negar la existencia del fenómeno.

El supuesto de la coexistencia de creencias contradictorias ha sido cuestionado de manera convincente por muchos estudios filosóficos (Mele, 1987, 2001; Talbot, 1995; Barnes, 1997). Como señalamos en la sección precedente, es pertinente observar que las creencias contradictorias podrían no coexistir, sino suceder una a la otra en el tiempo. Sin embargo, una alternativa más radical consiste en negar que la creencia verdadera forme parte inicialmente del repertorio de creencias del agente y que luego sea sustituida por la creencia falsa; esto es, las creencias contradictorias no coexisten ni se suceden temporalmente en la mente del agente. En Mele (2001), por ejemplo, puede encontrarse incluso un modelo explicativo del autoengaño que prescinde de tal supuesto y basado en sólida evidencia empírica provista por varios estudios psicológicos. En esta perspectiva el autoengaño es el producto de una combinación de sesgos cognitivos “fríos” y estados motivacionales y emocionales, que conduce a la adopción de creencias falsas sin que la creencia verdadera o más fundada en la evidencia haya estado alguna vez en posesión del agente. Huelga aclarar que la pregunta relativa a la coexistencia de creencias contradictorias permanece como problema no resuelto, por lo que sería un error considerar que estos estudios saldan de modo definitivo el debate. Sin embargo,

indudablemente constituye un error el suponer sin más que tal rasgo caracteriza esencialmente al autoengaño, soslayando los múltiples argumentos que se han opuesto a esta tesis. Como señalamos al inicio de este apartado, en consecuencia, el descuido de precisiones conceptuales torna dudoso que las investigaciones empíricas basadas en caracterizaciones tan cuestionables constituyan avances genuinos en nuestra comprensión del fenómeno del autoengaño.

IV

Encontramos, entonces, que el descuido de análisis conceptuales detallados (muchos de los cuales provienen de la filosofía) conduce a consecuencias teóricas y empíricas inadecuadas o insuficientemente fundadas. Ahora bien, como adelantamos, existe una segunda razón para considerar la posible transición del problema como un proceso complejo e inacabado, en el que hay tanto ganancias como pérdidas, o al menos problemas nuevos. Esta segunda razón se basa en el hecho de que parecería que nos encontramos lejos de una teoría explicativa unificada del fenómeno lo que, a su vez, nos conduce a nuevos interrogantes filosóficos. Una enumeración comprensiva (aunque no necesariamente exhaustiva) de las explicaciones del autoengaño incluye:

- a. Explicaciones neuropsicológicas (Ramachandran, 1996, 1998; Hirstein, 2000). El común denominador de este tipo de explicaciones es la comprensión de la patología neurológica, expresada en déficits como la anosognosia o la confabulación, como la base para la explicación del funcionamiento mental en personas en quienes pueden observarse procesos de autoengaño no patológico.
- b. Explicaciones evolucionistas (Van Leeuwen, 2007b, 2008; von Hippel y Trivers, 2011). Estas explicaciones se caracterizan o bien por postular una función evolutiva para el autoengaño (esto es, una contribución a la capacidad reproductiva del organismo), o bien por concebir este fenómeno como un subproducto estructural (*spandrel*) de sistemas seleccionados por su contribución a la aptitud.
- c. Explicaciones “disolucionistas” (Lewis, 1996; Clegg y Moissinac, 2005). Para estas perspectivas, basadas en filosofías construccionistas y relativistas, el autoengaño no existe dentro del yo individual del así llamado “autoengañado”. Existe sólo en la mente (más precisamente en las narrativas) del observador, quien

supone acceder a información más completa o correcta o a una representación “verdadera” del mundo. El autoengaño es un fenómeno cultural, no natural.

- d. Explicaciones cognitivo-motivacionales (Lazar, 1999; Mele, 1997, 2001). Para esta perspectiva el autoengaño es el resultado de una compleja combinación de procesos sesgados de procesamiento cognitivo y emotivo-motivacionales, combinación en la que los deseos del agente suelen jugar un rol decisivo en la producción del fenómeno.
- e. Explicaciones psicológico-sociales (Cohen, 2001; Zeruvabel, 2006). Para esta perspectiva, el autoengaño (o, como lo denominan, “negación” o “conspiración de silencio”) es el resultado de múltiples procesos y mecanismos tanto individuales como sociales, no un proceso unitario en el interior de un individuo. El autoengaño se descompone en una multiplicidad de categorías, discriminadas a través de criterios tales como su status psicológico, su relación con el individuo o el colectivo y su ubicación temporal, entre otros.⁵

⁵ Los árbitros de *Páginas de Filosofía* han señalado que, si bien se entiende que la enumeración de las posibles explicaciones del autoengaño no puede ser exhaustiva, se echa de menos una justificación desde el punto de vista de su base empírica, por una parte, y desde su rigor lógico, por la otra. Respecto de esta cuestión entiendo que no es posible proporcionar una justificación lógica de esta clasificación de las explicaciones, si por tal se entiende una distinción debida a características “formales” de las explicaciones, esto es, por el tipo de implicación entre los enunciados que reflejan las regularidades a explicar y aquellos que los explican (por ejemplo, relaciones deductivas o no deductivas), o la naturaleza de los enunciados legaliformes que fundamentan las explicaciones, en caso de que los hubiera. Sí entiendo que la clasificación presentada puede ser justificada, al menos parcialmente, a partir dos criterios diferentes relacionados con su base empírica. El primer criterio se basa en la naturaleza “próxima” o “distal” de las explicaciones. De este modo, es posible distinguir entre explicaciones que intentan proveer una comprensión de los mecanismos que hacen posible el autoengaño tal como podemos atribuirlo actualmente, de aquellas que procuran proveer de una comprensión de las causas de su origen en nuestra especie a lo largo del proceso evolutivo. El segundo criterio refiere a los niveles o estratos ontológicos a los que se apela para explicar el autoengaño. Este criterio permitiría distinguir distintos tipos de explicación del fenómeno (no necesariamente excluyentes), sobre la base de los procesos psicológicos, sociales o neurales que permitirían comprender los mecanismos de su producción. Esta clasificación es menos que satisfactoria como mínimo por dos razones. En primer lugar, las explicaciones “disolucionistas” (que merecerían ser llamadas también “eliminativistas”, dado que niegan la existencia del autoengaño) no se acomodan de modo aceptable en ninguna de las clases descriptas. Por otro lado, una

Pese a que esta multiplicidad de explicaciones podría ser evaluada, *prima facie*, como un avance en nuestro conocimiento del autoengaño, un análisis algo más sutil revela que está lejos de ser obvio que tal proliferación explicativa constituya un avance epistémico; por el contrario, podría ser vista no como una ventaja, sino como un demérito: demasiadas explicaciones de una misma clase de hechos conspirarían contra nuestro conocimiento.

Valga como ejemplo al respecto el planteo de J. Kim acerca de la coexistencia de explicaciones acerca de un mismo fenómeno. En diversos trabajos Kim ha defendido un principio general relativo a la explicación científica, principio denominado de la “exclusión explicativa”. Ha distinguido dos versiones diferentes de este principio: el principio *metafísico* de exclusión sostiene que dos explicaciones diferentes de un mismo *explanandum* “pueden ser ambas explicaciones correctas sólo si al menos alguna de las dos es incompleta o una es dependiente de la otra” (1989, p. 257); el principio *epistemológico* sostiene que nadie puede aceptar dos explicaciones “a menos que tenga una explicación apropiada de cómo están relacionadas entre sí” (*ibid.*, p. 257). Kim agrega que, en su versión metafísica, el principio puede aceptarse solo si a la vez se acepta alguna forma de *realismo explicativo*, ya que las nociones de incompletud o dependencia implican alguna relación objetiva en el mundo. Pero, observa, en su versión epistemológica el principio es aceptable aun cuando se renuncie a la idea de que existe alguna relación explicativa objetiva en el mundo, ya que puede encontrarse perturbador y disonante el hecho de tener que tratar con dos (o más) explicaciones independientes del mismo fenómeno. No es necesario, considera Kim, que los respectivos *explanans* sean lógicamente contradictorios, e incluso pueden tenerse suficientes garantías en favor de la verdad de cada conjunto de premisas. Sin embargo, aceptar ambos podría inducir a algún tipo de incoherencia en nuestro sistema de creencias. De este modo, “*demasiadas explicaciones pueden ser una fuente de incoherencia en vez de incrementar la coherencia*” (*ibid.*, p. 258. Cursivas del autor).

clasificación de explicaciones que apele simultáneamente a dos criterios sólo puede ser parcialmente satisfactoria desde la perspectiva de los requisitos formales exigibles a las clasificaciones. No obstante, como parecen dictar normas básicas de racionalidad, es preferible una clasificación no enteramente satisfactoria a la carencia absoluta de un esquema clasificatorio.

Sobre la base de lo anterior, es posible plantear si la proliferación de explicaciones sobre el autoengaño puede constituir un caso de sobreabundancia explicativa y, con ello, el problema de dar cuenta cómo tales explicaciones se relacionan entre sí. Pueden sugerirse algunas respuestas a este problema. Por un lado, es posible postular que los *explanans* de estas explicaciones no son los mismos, por lo que no compiten entre sí; simplemente estaríamos antes un caso de coexistencia explicativa basada en los postulados teóricos de distintas disciplinas. Sin embargo, esta respuesta requiere de mayor elaboración, ya que no es obvio que los distintos *explanans* sean diferentes. Por caso, las explicaciones “disolucionistas” y las cognitivo-motivacionales parecen estar claramente en competencia o conflicto; mientras que las primeras reducen el autoengaño a un fenómeno cultural, no intrapsíquico, que no tiene existencia en la mente de un agente aislado, las segundas lo reconocen como un fenómeno que ocurre en la mente de un sujeto, que puede describirse y explicarse con relativa independencia de factores culturales o sociales. Si bien puede pensarse una posible compatibilidad entre los dos tipos de explicaciones (al menos con versiones “moderadas” del construccionismo social), esta alternativa requiere de una elaboración conceptual importante, que hasta donde sabemos no ha sido realizada; la coexistencia explicativa continúa, por lo tanto, en estado de problema.

V

Como hemos intentado mostrar, el desarrollo de los estudios sobre el autoengaño en las últimas décadas ha implicado un incremento tanto de la cantidad y variedad de preguntas formuladas como de los desarrollos teóricos sobre el problema. No obstante, como creo que se pone de manifiesto a través de las dos secciones precedentes, tales desarrollos también han implicado el surgimiento de problemas conceptuales y epistemológicos nuevos.

En esta sección final describiré una línea de investigación actual y plausiblemente promisoría para resolver al menos algunos de los enigmas conceptuales del autoengaño. Como se recordará, en la primera sección presentamos una caracterización “de trabajo” de este fenómeno, y señalamos que todos sus componentes pueden ser objetados. Este cuestionamiento puede ser tan radical como para poner en tela de juicio uno de los supuestos más profundamente arraigados acerca del autoengaño, esto es, la tesis según la cual el producto de un proceso de autoengaño debe ser necesariamente una creencia. Las posiciones que

describiré a continuación comparten justamente, y *grosso modo*, la objeción a la primacía de la creencia como componente esencial del autoengaño.

Pese a que, como adelantamos, tal línea de investigación constituye una vertiente actual, la estrategia consistente en cuestionar la primacía de la creencia como resultado de un proceso de autoengaño no es enteramente nueva. Uno de los intentos de explicar ciertos aspectos intrigantes del autoengaño se ha basado en la estrategia de negar que el producto de este fenómeno sea una creencia; en vez de ella, el resultado del autoengaño sería una *manifestación* [*avowal*], tesis sostenida por Audi (1982) y por Rey (1988). Para ambos autores una manifestación es básicamente una disposición a afirmar una proposición con “sinceridad”; sin embargo, tal proposición carece de conexiones profundas con la acción. Entonces, si quien se autoengaña meramente manifiesta que *p* encontrándose en el estado de autoengaño de que *p*, no sería necesario atribuir al agente tanto la creencia de que *p* como la creencia de que no *p*, con lo cual se evitaría el problema de explicar la coexistencia de creencias contradictorias. Cuando un agente está autoengañado respecto de que *p*, tanto Audi como Rey sostienen que el agente posee una creencia (que Rey llama “creencia central”) de que no *p*. Como ha observado Van Leeuwen (2007a) la disociación entre creencias y acciones es un componente esencial del enfoque de la manifestación; de otro modo resultaría completamente oscuro en qué respecto se supone que una manifestación es distinta de una creencia genuina. Para Rey, podría decirse que, *ceteris paribus*, una persona cree manifiestamente que *p* si afirmara sincera y decididamente que *p* si se le preguntara al respecto. Creencias manifestadas, para este autor, son aquellas que se atribuyen sobre la base de la conducta verbal, mientras que creencias centrales son aquellas que se atribuyen sobre la base de las acciones. Audi, a su vez, advierte que el hecho que S esté autoengañado con respecto a *p* no implica su creencia –incluso consciente– de que *p*; lo que su posición requiere respecto de la actitud positiva de S hacia *p* es que S esté sinceramente dispuesto a manifestarlo.⁶

Ahora bien, dentro de la línea de negación de la creencia dos alternativas más radicales y novedosas que describiremos brevemente

⁶ Como otros debates filosóficos acerca del autoengaño, la viabilidad del enfoque de la manifestación permanece como una cuestión abierta. Para un examen acerca de si el resultado del autoengaño puede ser una manifestación, y no una creencia, cfr. Van Leeuwen (2007a).

son la que postulan estados intencionales intermedios y alternativos entre la creencia y deseos, y las posiciones denominadas “no doxasticistas” acerca del autoengaño, que se caracterizan por negar que la creencia sea necesaria para comprender el fenómeno.

Un análisis representativo de la primera de estas posiciones puede hallarse en Egan (2008). En un examen de los posibles vínculos entre el autoengaño y el delirio sugiere que podría resultar útil, desde el punto de vista explicativo, postular una actitud intermedia entre dos de las actitudes proposicionales más familiares: el deseo y la creencia. Algunos casos de autoengaño, en su opinión, manifiestan el tipo de aislamiento e insensibilidad ante la evidencia que es característico de los delirios. En tales casos, prosigue, podemos tener cierta reticencia ante la afirmación de que quien se autoengaña cree genuinamente en la proposición en cuestión; sin embargo, tampoco parece correcto decir que el sujeto meramente desea que lo sea. En vez de ambas posibilidades, sugiere, parecería encontrarse en un estado intermedio entre la creencia y el deseo, un tipo de estado al que denomina “creseo” [*besire*]. Tal tipo de estado haría posible para la “creencia” de quien se autoengaña ser insensible a la evidencia que muestra su falsedad, y sin embargo jugar en alguna medida el rol de la creencia en la guía de la conducta; también haría posible explicar los casos en los cuales la “creencia” de quien se autoengaña tiene un rol empobrecido tanto en la guía de la conducta como en términos inferenciales, como puede observarse en algunos casos. Egan aclara que no pretende sugerir que la postulación de actitudes intermedias como la descrita constituya la explicación correcta del autoengaño en general, ya que este fenómeno es multifacético y no resultará posible encontrar una concepción explicativa general y unificada. Lo que quiere sugerir es que este tipo de estado intermedio es el camino correcto para describir lo que ocurre en alguna clase acotada de casos que pueden ser plausiblemente agrupados bajo el rótulo “autoengaño”, esto es, aquellos que expresan el tipo particular de independencia y aislamiento de la evidencia que puede observarse en los delirios.

Por último, puede hallarse una clara exposición de las posiciones no doxasticistas respecto del autoengaño en Edwards (2013). Esta autora examina el problema al que denomina “problema doxástico” del autoengaño, el cual requiere que se especifique si el estado de autoengaño respecto de p implica que el sujeto cree simultáneamente que p y que no p , que posee una u otra creencia o, como ella sostendrá, no posee ninguna creencia acerca de p . Esta posición, a la que considera

ampliamente soslayada dentro de los estudios sobre el fenómeno (a la que denomina, justamente, “no doxasticismo”) sostiene que la persona no posee la creencia de que p ni la creencia de que no p.⁷ Edwards considera que los requisitos explicativos que tradicionalmente han sido satisfechos recurriendo a las creencias pueden ser cómodamente satisfechos apelando a otras categorías psicológicas; la viabilidad del no doxasticismo depende del reconocimiento de la riqueza de la psicología folk. Un ejemplo de la aplicación de esta estrategia sería el siguiente. Un hombre se autoengaña respecto de la fidelidad de su esposa, creyendo que ella no está teniendo un affaire. Si bien los enfoques tradicionales respecto del autoengaño afirmarían que el hombre posee como mínimo alguna creencia, Edwards considera que puede darse una explicación de su estado sin recurrir a tal concepto. De este modo, el hombre *sospecha* que su esposa puede estar teniendo un amorío e, inmediatamente, erige defensas contra esa sospecha. *Desea* fuertemente creer que su esposa le es fiel, y la combinación del deseo y la sospecha explican por qué comienza a evitar información que favorece la conclusión temida y a buscar pruebas contra ella, así como a distorsionar cualquier evidencia que encuentra. Aun así, experimenta inquietud respecto de esta cuestión; siente *ansiedad* respecto de la posibilidad de que su esposa esté teniendo un amorío, y tiene molestas *dudas* acerca de la información que ya posee, pero *espera* que ella no le sea infiel. No obstante, sus irritantes dudas y sospechas evitan que adquiera la creencia de que su esposa no le es infiel.

Tanto la postulación de entidades intermedias entre la creencia y el deseo como la posición no doxasticista constituyen alternativas más radicales ante algunos de los principales problemas planteados por el autoengaño. No hay ninguna razón concluyente, que sepamos, para considerar que estas alternativas son inviables. Sin embargo, conviene observar que se basan en la revisión de un supuesto muy arraigado acerca del autoengaño y aceptado por una amplia mayoría de especialistas, como es el supuesto relativo al carácter privilegiado de las creencias dentro de las actitudes proposicionales, por lo que resta por evaluar si las ganancias explicativas derivadas de su rechazo compensan las seguras pérdidas que implica.

En cualquier caso, sin duda asistimos a un complejo y notable proceso por el cual un problema tradicionalmente filosófico se convierte

⁷ Variantes de la posición no doxasticista pueden hallarse en Funkhouser (2009) y Porcher (2012).

en un problema científico (al menos parcialmente); los desarrollos de los próximos años nos permitirán evaluar con más claridad si los estudios sobre el autoengaño derivan hacia una perspectiva científica unificada o, empleando las célebres palabras de John Austin, “un planeta frío y bien regulado, que progresa sin pausa hacia un distante estado final” (1961, pp. 180).

Bibliografía:

- Adler, J. E. (2002), *Belief's Own Ethics*, Cambridge, The MIT Press.
- Asociación Psiquiátrica Americana (1994), *Manual Diagnóstico y Estadístico de los Trastornos Mentales, Cuarta Edición*, Barcelona, Masson.
- Audi, R. (1982), “Self-Deception, Action, and Will”, *Erkenntnis* 18, pp. 133–58.
- Austin, J. (1961), *Philosophical Papers*, editado por G. J. Warnock y J. O. Urmson, Oxford, Clarendon Press. Citado por Genaro Carrió y Eduardo Rabossi, “La filosofía de John L. Austin”, en John Austin, *Cómo hacer cosas con palabras*, Barcelona, Paidós, 1990, p. 27.
- Bach, K. (1981), “An Analysis of Self Deception”, *Philosophy and Phenomenological Research*, 41, 3, pp. 351-370.
- Badhwar, N. (2008), “Is Realism Really bad for You? A Realistic Response”, *The Journal of Philosophy*, CV, 2, pp. 85-107.
- Bandura, A. (1991), “Social cognitive theory of moral thought and action”, en W. M. Kurtines & J. L. Gewirtz (eds.), *Handbook of moral behavior and development. Vol. 1*. Hillsdale, NJ, Erlbaum, pp. 45-103.
- Bandura, A. (2011), “Self-deception: A paradox revisited”, *Behavioral and Brain Sciences*, 34, pp. 16-17.
- Barnes, A. (1997), *Seeing through self-deception*, New York, Cambridge University Press.
- Baron, M. (1988), “What is Wrong with Self-Deception?” en Brian McLaughlin & Amelie Oksenberg-Rorty (eds.), pp. 434-449.
- Bennett, J. (1990), “Why Is Belief Involuntary?” *Analysis*, 50, 2, pp. 87-107.
- Bermúdez, J. L. (2000), “Self Deception, intentions and contradictory beliefs”, *Analysis* 60, 4, pp. 309-319.

- Bok, S. (1980), “The self deceived”, *Social Science Information* 19, pp. 923-935.
- Borge, S. (2003), “The Myth of Self-Deception”, *The Southern Journal of Philosophy*. 41, 1, pp. 1–28.
- Butler, J. (1726), “Sermon X. Upon Self-Deceit”, recuperado el 24/07/2010 de <http://anglicanhistory.org/butler/rolls/10.html>
- Clegg, J. W. & L. Moissinac (2005), “A relational theory of self-deception”, *New Ideas in Psychology*, 23, pp. 96–110.
- Cohen, S. (2001), *States of Denial. Knowing about Atrocities and Suffering*, Oxford, Blackwell.
- Colvin, R. & J. Block (1994), “Do Positive Illusions Foster Mental Health? An Examination of the Taylor and Brown Formulation”, *Psychological Bulletin*, 116, 1, pp. 3-20.
- Davidson, D. (1985), “Engaño y división”, en *Mente, mundo y acción*, Barcelona, Paidós, pp. 99-117.
- Demos, R. (1960), “Lying to Oneself”, *Journal of Philosophy*, 57, pp. 588–95.
- Deweese-Boyd, I. (2008), “Collective self-deception, collective injustice: Consumption, sustainability and responsibility”, recuperado el 27/10/10 de http://www.colorado.edu/philosophy/center/rome/papers/DeWeese-Boyd_CollectiveSelfDeception_CollectiveInjustice.pdf
- Edwards, S. (2013), “Nondoxasticism about Self-Deception”, *Dialectica* Vol. 67, N° 3, pp. 265–282.
- Egan, A. (2008), “Imagination, Delusion, and Self-Deception”, en T. Bayne and J. Fernandez (eds.) *Delusion and Self-Deception: Motivational and Affective Influences on Belief-Formation*, Hove, East Sussex, Psychology Press.
- Erez, A., D. E. Johnson & T. A. Judge (1995), “Self-Deception as a Mediator of the Relationship Between Dispositions and Subjective Well-Being”, *Personality and Individual Differences*. 19. 5, pp. 597-612.
- Fingarette, H. (1969), *Self-deception*, London, Routledge & Kegan Paul.
- Funkhouser, E. (2005), “Do the Self Deceived Get what they Want?” *Pacific Philosophical Quarterly*. 86, pp. 295 –312.
- Funkhouser, E. (2009), “Self-Deception and the Limits of Folk Psychology”, *Social Theory and Practice*, 35, pp. 1–13.

- Gur, R. & H. Sackeim (1979), “Self-Deception: A Concept in Search of a Phenomenon”, *Journal of Personality and Social Psychology* 37, pp. 147–69.
- Harré, R. (1988), “The Social Context of Self-Deception”, en Brian McLaughlin & Amelie Oksenberg-Rorty (eds.), pp. 164-379.
- Hirstein, W. (2000), “Self-Deception and Confabulation”, *Philosophy of Science*. 67, pp. 418-429.
- Johnston, M. (1995), “Self-deception and the nature of mind”, en McLaughlin, Brian y Amelie Oksenberg-Rorty (eds.) (1988), *Perspectives on Self-Deception*, Berkeley, University of California Press, pp. 63-91.
- Kim, J. (1989), “Mecanismo, propósito y exclusión explicativa”, *Análisis Filosófico*, X (1), pp. 15-47.
- Kipp, D. (1980), “On Self-Deception”, *Philosophical Quarterly* 30, pp. 305–17.
- Kurzban, R. & C. A. Aktipis (2007), “Modularity and the social mind: Are psychologists too selfish?” *Personality and Social Psychology Review*, 20, pp. 1–19.
- Lazar, A. (1999), “Deceiving Oneself or Self-Deceived?” On the Formation of Beliefs ‘Under the Influence’, *Mind* 108, 430, pp. 265-290.
- Levy, N. (2004), “Self-deception and Moral Responsibility”, *Ratio (new series)*, XVII, pp. 294-311.
- Lewis, B. (1996), “Self-deception: A postmodern reflection”, *Journal of Theoretical and Philosophical Psychology*, 16, 1, pp. 49-66.
- Martin, M. (2012), *Happiness and the Good Life*, New York, Oxford University Press.
- McLaughlin, B. & A. Oksenberg-Rorty (eds.) (1988), *Perspectives on Self-Deception*, Berkeley, University of California Press, pp. 29-62.
- Mele, A. (1987), *Irrationality. An Essay on Akrasia, Self-Deception, and Self-Control*, New York-Oxford, Oxford University Press.
- Mele, A. (1997), “Real Self Deception”, *Behavioral and Brain Sciences* 20 (1), pp. 91-136.
- Mele, A. (2001), *Self-deception Unmasked*, Princeton, Princeton University Press.

- Mele, A. (2003), “Emotion and Desire in Self-Deception”, en Anthony Hatzimoysis (ed), *Philosophy and the Emotions*, Cambridge, Cambridge University Press, pp. 163-179.
- Oksenberg-Rorty, A. (1988), “The Deceptive Self: Liars, Layers and Lairs”, en Brian McLaughlin & Amelie Oksenberg-Rorty (eds.), pp. 11-28.
- Paluch, S. (1967), “Self-Deception”, *Inquiry* 10, pp. 268–78.
- Paulhus, D. (2007), “Self Deception”, en Roy Baumeister & Kathleen Vohs (eds.). *Encyclopedia of Social Psychology*. Los Angeles, Sage. pp. 802-804.
- Pears, D. (1991), “Self-Deceptive Belief-Formation”, *Synthese* 89, pp. 393–405.
- Porcher, J. (2012), “Against the Deflationary Account of Self-Deception”, *Humana.Mente. Journal of Philosophical Studies*, 20, pp. 67–84.
- Ramachandran, V. S. & S. Blakeslee (1998), *Phantoms in the Brain. Probing the Mysteries of the Human Mind*, New York, Harper Collins.
- Ramachandran, V. S. (1996), “The Evolutionary Biology of Self-Deception, Laughter, Dreaming and Depression: Some Clues for Anosognosia”, *Medical Hypothesis* 47, pp. 347-362.
- Rey, G. (1988), “Toward a computational account of Akrasia and self-deception”, en A. O. Rorty & B. P. McLaughlin (eds.), pp. 264-296.
- Ryan, S. (2003), “Doxastic compatibilism and the ethics of belief”, *Philosophical Studies*, 114, pp. 47–79.
- Scott-Kakures, D. (2000), “Motivated Believing: Wishful and Unwelcome”, *Nous* 34, 3, pp. 348–375.
- Scott-Kakures, D. (2001), “High anxiety: Barnes on what moves the unwelcome believer”, *Philosophical Psychology*, 14, 3, pp. 313-326.
- Scott-Kakures, D. (2009), “Unsettling Questions: Cognitive Dissonance in Self-Deception”, *Social Theory and Practice*, 35, 1, pp. 73-106.
- Sharma, M., Modi, S., Khushu, S., & Mandal, M. K. (2010), “Neural activation pattern in self-deceivers”, *Psychological Studies*, 55(1), pp. 71–76.
- Steup, M. (2008), “Doxastic freedom”, *Synthese*, 161, pp. 375–392.

- Surbey, M. (2011), “Adaptive significance of low levels of self-deception and cooperation in depression”, *Evolution and Human Behavior* 32, pp. 29–40.
- Talbott, W. (1995), “Intentional Self-Deception in a Single Coherent Self”, *Philosophy and Phenomenological Research* 55, 1, pp. 27-74.
- Talbott, W. J. (1995), “Intentional Self-Deception in a Single Coherent Self”, *Philosophy and Phenomenological Research* 55, 1, pp. 27-74.
- Taylor, S. & J. Brown (1988), “Illusion and Well-Being: A Social Psychological Perspective on Mental Health”, *Psychological Bulletin* 103, 2, pp. 193-210.
- Trivers, R. (2000), “The Elements of a Scientific Theory of Self-Deception”, *Annals of the New York Academy of Sciences* 907, pp. 114-131.
- Trivers, R. (2010), “Deceit and Self-Deception”, en Peter M. Kappeler & Joan B. Silk (eds.), *Mind the Gap. Tracing the Origins of Human Universals*, Springer, pp. 373-393
- Van Leeuwen, D. S. N. (2007a), “The Product of Self-Deception”, *Erkenntnis* 67, pp. 419–437.
- Van Leeuwen, D. S. N. (2007b), “The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon”, *Philosophical Psychology* 20, 3, pp. 329–348.
- Van Leeuwen, D. S. N. (2008), “Finite rational self-deceivers”, *Philosophical Studies* 139, pp. 191–208.
- Van Leeuwen, D. S. N. (2009), “Self-Deception Won’t Make You Happy”, *Social Theory and Practice*, 35, 1, pp. 107-132.
- von Hippel, W. & R. Trivers (2011), “The evolution and psychology of self-deception”, *Behavioral and Brain Sciences* 34, pp. 1–56.
- Williams, B. (1973), “Deciding to believe”, en *Problems of the Self. Philosophical Papers 1956-1972*, Cambridge, Cambridge University Press, pp. 136-151.
- Zerubavel, E. (2006), *The Elephant in the Room. Silence and Denial in Everyday Life*, New York, Oxford University Press.

Recibido el 24 de marzo de 2018; aceptado el 05 de noviembre de 2018.