

teorema

Vol. XXVI/3, 2007, pp. 19-30

ISSN: 0210 - 1602

La etiología del autoengaño. ¿Pretendo engañarme o me engañan mis mecanismos?

M. Rosario Hernández Borges

ABSTRACT

Most philosophical accounts of self-deception are in one of two groups: the intentional explanation, which considers intention as a basic element to explain the process of self-deception; or, the non-intentional explanation, which explains self-deception as the result of motivational bias. I will take Davidson as a representative author of the first model, and Mele of the second. In this paper I will argue that the intentional model does not explain self-deception successfully; and that Mele's non-intentional explanation has to be completed to avoid the "selective problem" that Bermúdez has formulated. Three elements have to be considered to complete the etiology of self-deception: emotion, coherence and self-knowledge.

RESUMEN

La mayoría de las explicaciones filosóficas del autoengaño son de carácter intencional, en tanto consideran la intención como el elemento básico para explicar el proceso del autoengaño, o son no intencionales, y explican el autoengaño como el resultado de un sesgo motivacional. Consideraré a Davidson como un autor representativo del primer grupo, y a Mele como un ejemplo del segundo. En este trabajo sostendré que el modelo intencional no explica adecuadamente el autoengaño; y que la explicación no intencional de Mele ha de ser completada para evitar "el problema selectivo" que Bermúdez plantea. Para ello hemos de considerar tres nuevos elementos: la emoción, la coherencia y el autoconocimiento.

El autoengaño es un fenómeno muy familiar para todos. En ocasiones nos damos cuenta de que en nuestros juicios o en la planificación de nuestras acciones no hemos tenido en cuenta hechos importantes, no meramente por olvido, sino como si "no hubiéramos querido verlos". Si no podemos saber que nos estamos autoengañando mientras el autoengaño está sucediendo, esto podría ser una razón para pensar que el autoengaño es un proceso inconsciente. Sin embargo, el acceso directo y privilegiado a los estados mentales propios tiene un lugar tan prioritario en nuestro pensamiento y acción que nos lleva a considerarlos erróneamente como la suma de nuestras causas y razones para actuar, olvidando que son el resultado de procesos mentales y sólo,

en algunos casos, juegan un papel causal consciente. Reducir nuestra vida mental a un proceso consciente de evaluación racional de creencias y deseos hace del autoengaño uno de los fenómenos mentales más difícil de explicar. La explicación que da Davidson del autoengaño es un buen ejemplo de estas dificultades (sección I). Las explicaciones no intencionales, como la de Mele, que consideran el autoengaño como el resultado de sesgos motivacionales se han propuesto como alternativa para evitar paradojas (sección II). Sin embargo, el sesgo motivacional no parece suficiente para dar una explicación general de cualquier caso de autoengaño. Por el contrario, como Bermúdez señala, el autoengaño es paradigmáticamente selectivo (sección III). Con el fin de lograr un modelo más completo del autoengaño y evitar el “problema selectivo”, propondré que además de la no intencionalidad y del sesgo motivacional, es necesario considerar tres factores más: la emoción, la coherencia y el autoconocimiento (sección IV).

I. AUTOENGAÑO CON CONCIENCIA. LA EXPLICACIÓN DE DAVIDSON.

En “Deception and Division” (1986), Davidson caracteriza el autoengaño como la situación habitual donde “la suma de la evidencia disponible para el agente indica la verdad de una proposición, lo que inclina al agente a creerla (hace que el agente la trate como si fuera más probable su verdad que su falsedad). *Esta inclinación* (la alta probabilidad subjetiva) *causa que él [...] busque, favorezca o enfatice la evidencia a favor de la falsedad de la proposición, o ignore la evidencia a favor de su verdad*” [Davidson (1986), p. 80 (La cursiva es mía)].

Siguiendo a Davidson, las condiciones bajo las cuales un agente se autoengaña son las siguientes: (i) El agente tiene evidencia sobre cuya base cree que es más acertada la verdad de p que su negación. (ii) El pensamiento de que p motiva al agente a actuar a fin de causar la creencia en la negación de p . (iii) La acción implicada puede consistir en apartar intencionalmente la atención lejos de la evidencia a favor de p , o puede suponer la búsqueda activa de evidencia contra p . (iv) El motivo se origina en una creencia de que p es verdadera. (v) La acción se realiza con la intención de producir una creencia en la negación de p . (vi) El estado que motiva el autoengaño y el estado que lo produce coexisten. La creencia de que p no sólo causa una creencia en la negación de p , sino que también la sostiene.

Hay muchos ejemplos habituales. La mujer que se autoengaña acerca de las acciones de su marido es uno de los casos más comunes. Su marido llega más tarde que antes, no está interesado en lo que sucede en la vida de su mujer, empieza a hacer viajes de negocios algunos fines de semana. Ella tiene evidencia suficiente como para creer en la infidelidad de su marido (condición i). Es precisamente esta creencia lo que motiva que ella niegue la verdad

acerca de ese pensamiento (condiciones ii y iv). Por tanto, la mujer tiene que hacer un esfuerzo intencional para olvidar la evidencia o buscar evidencia contra la infidelidad de su marido (condición iii). La mujer actúa de esta manera con el fin de creer que no es verdad que su marido le sea infiel (condición v). En conclusión, la mujer cree tanto en la infidelidad de su marido como en la fidelidad de su marido porque la primera creencia es necesaria para causar y mantener la segunda creencia.

Esta explicación podría parecer extraña porque presupone que el que se autoengaña mantiene creencias contradictorias a la vez. La paradoja es que cuando se interpreta al sujeto que se autoengaña no podemos decir que crea y no crea la misma proposición, porque, de otra forma, como intérpretes caeríamos en contradicciones al describir la confusión del que se autoengaña. El autoengaño es un problema para el intérprete porque para entender a un agente es necesario identificar un alto grado de consistencia en sus creencias, y entre creencias y acciones, lo que implica racionalizar las creencias y las acciones. Sin embargo, aunque el que se autoengaña tiene una razón para librarse de la creencia que encuentra desagradable o dolorosa o que lo desacredita, según Davidson, ésta no es una razón a favor de la creencia. Pueden darse muchas causas para explicar una acción, pero no todas serían razones a favor de una creencia o de una acción. Únicamente las razones (o causas) que son la conclusión lógica a partir del sistema de creencias del agente se considerarían como razones para creer o actuar. “Si estuviera más feliz, más orgulloso, más relajado, menos nervioso, si tuviera una cierta creencia, esa es una razón, dejando a un lado otras consideraciones, para tener la creencia. Pero tal razón no es, en sí misma, una razón para suponer que la creencia deseable es verdadera” [Davidson (1998), p. 4]. Lo que Davidson afirma es que, en ese caso, faltaría la conexión lógica entre las razones que damos para explicar las creencias y acciones del agente y esas mismas creencias y acciones.

Así que, ¿cómo podemos explicar el autoengaño? Davidson propone dos formas de resolver el problema:

1. La primera solución admitiría “una forma impura de explicación que [...] es causal [...] pero no racionaliza lo que explica” [Davidson (1998), p. 7]. En opinión de Davidson, si los sucesos mentales se describen únicamente en términos causales, como fuerzas ciegas que gobiernan las propiedades físicas o fisiológicas, su naturaleza racional o irracional desaparece. La conducta física se explica en términos físicos, pero las acciones intencionales se ven como parte de una pauta racional; y el autoengaño es difícil de explicar porque tendemos a considerar al agente como racional. Por ello, si damos una explicación causal, cambiamos el tema y la naturaleza racional o irracional de la acción desaparece. En palabras de Davidson: “El deseo de cambiar una creencia

causa el cambio, pero no es una razón para considerar verdadera a la nueva creencia y falsa a la antigua” [Davidson (1998), p. 8].

Si consideramos la naturaleza intencional de las acciones y sucesos mentales, entonces sus causas tendrán que ser también sus razones. Sin embargo, esto sugiere que el autoengaño se racionalizaría. Por tanto, tenemos un dilema: “Para una explicación de un efecto mental necesitamos una causa mental que sea también una razón a favor de ese efecto, pero, si la tenemos, el efecto no puede ser un caso de irracionalidad” [Davidson (1982), pp. 299-300]. En conclusión, esta solución no puede ser considerada seriamente.

2. La segunda solución, que ha recibido muchas críticas, se propone para evitar la contradicción de creer p y $\neg p$ simultáneamente. Davidson piensa que podemos tener dos creencias contrarias si éstas se mantienen separadas. La imagen que traza es la de una mente dividida, donde, metafóricamente, “un muro separaba las creencias que, si llegaran juntas a la conciencia, destruiría al menos a una” [Davidson (1998), p. 8]. En opinión de Davidson, todos compartiríamos un núcleo central compuesto por la mayoría de verdades comunes y diferiríamos en algunos detalles disonantes. La condición esencial de una explicación correcta es que el núcleo central muestre un grado más alto de coherencia y racionalidad que las otras partes.

Esta segunda solución no es clara en absoluto. En primer lugar, Davidson señala que, tomando a Platón, Aristóteles, san Agustín o Freud como ejemplos de autores que defendieron esta idea, p y $\neg p$ no pueden representarse juntas en la conciencia. Sin embargo, añade que las divisiones de la mente no son el resultado de la conciencia, el coraje, el intelecto, o el ello. Por tanto, no explica cómo tiene lugar esa división. Su propósito es “[...] distinguir estrictamente entre aceptar una proposición contradictoria y aceptar por separado dos proposiciones contradictorias, donde esto último exige, o quizás sólo expresa, la idea de que los pensamientos se mantuvieran separados” [Davidson (1998), p. 9]. Pero el tema difícil de resolver es cómo saber qué proposiciones son contradictorias si se consideran separadamente. Solamente evaluándolas como parte de un sistema es posible establecer su contradicción. Si no hay contradicción, no hay irracionalidad ni autoengaño. Así que, una vez más, la solución de Davidson disuelve el problema que trata de resolver.

II. EL AUTOENGAÑO COMO RESULTADO DE UN SESGO MOTIVACIONAL. LA EXPLICACIÓN NO INTENCIONAL DE MELE

Davidson hace las dos propuestas anteriores no sólo para explicar el autoengaño, sino también para evitar paradojas. Mele señala que las dos para-

dojas que surgen en el problema del autoengaño se originan al intentar explicarlo en los mismos términos en los que explicamos el engaño. O bien A cree que p es falso y hace que B crea que p es verdadero, o bien A hace todo lo posible para que B no descubra ni su estrategia ni sus planes de engaño. Si A se engaña a sí mismo y se trasladan estas dos explicaciones del engaño al caso del autoengaño, A cree que p es falso y verdadero a la vez (sería lo que Mele llama la “paradoja estática”), o A intenta ocultar a A su estrategia de engaño (la “paradoja de la estrategia”).

El problema en estos dos casos, según Mele, es que la intención se supone en el proceso del autoengaño. La creencia falsa del sujeto que se autoengaña se confunde con su acción intencional. El error habitual es confundir la no accidentalidad del autoengaño con el hecho de que el agente debe engañarse a sí mismo intencionalmente.

Mele [1987, 1998, 2001] considera que el autoengaño no es engaño intencional, y que el conocimiento sesgado común en el autoengaño es provocado y sostenido por estados motivacionales (deseos). La primera paradoja desaparece si consideramos que “un tratamiento de los datos influido por el deseo tiene como resultado que la persona no adquiera la creencia verdadera y que él o ella adquiera (o conserve) la creencia falsa” [Mele (1998), p. 46]. No habría buenas bases para sostener que el que se autoengaña posee simultáneamente dos creencias contradictorias. Algo similar sucede en la paradoja estratégica. El que se autoengaña no intenta engañarse a sí mismo y “podemos entender cómo funcionan [...] los procesos causales sin que el agente orqueste el proceso” [Mele (2001) p. 61]. Así las cosas, el asunto a dilucidar sería qué clase de mecanismo o proceso no intencional causa el autoengaño.

Mele [1998, 2001] ofrece cuatro condiciones que conjuntamente serían suficientes para “caer en el autoengaño al adquirir una creencia de que p” [Mele (2001), p. 50]: (i) La creencia de que p que S adquiere es falsa. (ii) S trata los datos aparentemente relevantes de una forma sesgada motivacionalmente. (iii) Este tratamiento sesgado no es una causa desviada de la adquisición de la creencia de que p por parte de S. (iv) El conjunto de datos que S posee en ese momento proporciona mayor garantía a favor de p que a favor de q.

Las condiciones i y iv necesitan ser totalmente aceptadas para entender el autoengaño. Las condiciones ii y iii no son tan obvias y requieren alguna discusión.

La tercera condición ha sido un tema central en los debates sobre la racionalidad y la explicación causal de las acciones, y tiene que ver con la necesidad de evitar la identificación de sucesos fortuitos como las causas verdaderas de la acción. Un ejemplo muy conocido es el caso del sobrino que conduce hacia la casa de su tío para matarle con una pistola, pero en su camino, debido a su nerviosismo, atropella a un peatón que resulta ser su tío [Chisholm (1966); Searle (1983)]. La discusión filosófica sobre este problema se ha centrado en el papel que la intencionalidad juega en estas cadenas causales des-

viadas. En ellas la accidentalidad se identifica con no intencionalidad, por lo que deben quedar fuera de cualquier explicación causal de la acción intencional. Sin embargo, en el modelo de Mele, la dificultad para explicar el autoengaño no yace en la intencionalidad de las acciones o creencias sino en que las creencias no se han adquirido por casualidad, por accidente. Esto es, en los casos de las cadenas causales desviadas la accidentalidad se confunde con la no intencionalidad. En el caso del autoengaño, el agente cree ‘no p’ de manera no intencional, pero no es casual que crea ‘no p’ cuando tiene más evidencia a favor de ‘p’. La no casualidad se explica en la segunda condición suficiente que estipula Mele.

La segunda condición suficiente para adquirir una creencia, y también el aspecto más importante en esta discusión, se refiere a qué causa que un dato sea tratado de forma motivacionalmente sesgada. Mele recurre a investigación empírica para explicar el sesgo motivacional [Friedrich (1993), Trope y Liberman (1996)]. Combina dos modelos, el de Friedrich y el de Trope & Liberman, para formar lo que llama el modelo Friedrich-Trope-Liberman (FTL). Friedrich argumenta que la detección y minimización del error crucial es el principio central en la comprobación de hipótesis cotidianas. Cuando el error potencial se vuelve destacable, se ponen en movimiento estrategias de comprobación automáticas, probablemente favorecidas por la presión selectiva evolucionista. Esto no significa que nos concentremos principalmente en los casos que confirman la hipótesis que estamos probando. Por el contrario, la motivación también influye en la atención que damos a la evidencia que permite falsar la hipótesis. Al aplicar este modelo al autoengaño en relación con la autoestima, Friedrich mantiene que “un candidato para el error fundamental que nos interesa es considerar como verdadero algo que conduce a criticarse erróneamente a uno mismo o a reducir la autoestima. Tales costes son muy salientes y se pagan inmediatamente en términos de malestar psicológico” [Friedrich (1993), pp. 314].

Esta aportación de Friedrich se completa con el concepto de “umbral de confianza” del modelo de Trope y Liberman. “La aceptación es la confianza mínima en la verdad de una hipótesis requerida antes de aceptarla, y no continuar probándola, y el umbral de rechazo es la confianza mínima en la falsedad de una hipótesis requerida antes de rechazarla y dejar de probarla” [Trope y Liberman (1996), p. 253]. Trope y Liberman afirman que los dos umbrales no son igualmente exigentes y dependen del coste de la falsa aceptación (o el falso rechazo) en relación al coste de la información. Mientras que el coste de la información está constituido por las fuentes y los esfuerzos necesarios para obtener información relevante, los costes de falsa aceptación y falso rechazo son medidas de importancia subjetivas para el individuo. Mele las interpreta como los costes “que sería razonable que un agente esperara tener, dados los deseos y creencias del agente, si tuviera expectativas sobre cosas de ese tipo” [Mele (2001), p. 35]. Por tanto, los costes son relativos a la

propia situación motivacional del individuo. La pregunta, entonces, sería cómo los costes, las condiciones motivacionales, causan el autoengaño. La respuesta se encuentra en dos aspectos: cómo llegamos a tener creencias y cómo probamos las hipótesis.

Para ilustrar cómo el modelo FTL puede explicar la creencia resultado del autoengaño, Mele utiliza el ejemplo de Bob, que desea que sea verdad que él fuera el tercer mejor base en su liga. Bob analiza la estadística de la competición y se compara con otro base, Carl. A pesar de los datos, que muestran que Carl es mejor jugador que Bob, Bob concluye que él es mejor. El modelo FTL explica este caso señalando que Bob “tiene un umbral más bajo para creer que es mejor que Carl que para creer que no lo es, y esto a su vez se explica en gran parte por su deseo de que realmente él sea mejor jugador” [Mele (2001), p. 38]. Así que el soporte de evidencia necesario para obtener la creencia depende del umbral, que, a su vez, depende del deseo. Pero, ¿de qué depende el deseo? Volveré sobre este punto en la próxima sección. Ahora, el asunto es que el modelo FTL no requiere la intervención de la intención.

En la mayoría de los casos de autoengaño, están implicados procesos erróneos de comprobación de hipótesis. De acuerdo con el modelo FTL, la gente prueba las hipótesis minimizando el error costoso (entendido el coste en términos de malestar psicológico). Por tanto, se sigue que, en los casos relevantes, “debido en parte a la posición ocupada por un deseo de que p en la economía motivacional del agente *en ese momento*, la creencia falsa de que $\neg p$ sería un error más costoso *en ese momento* que la creencia falsa de que p ” [Mele (2001), p. 40, la cursiva es mía]. La evitación o minimización del error costoso propuesto por Friedrich junto con el concepto de umbral de confianza para la aceptación y el rechazo de proposiciones propuesto por Trope y Liberman explican el efecto que el deseo tiene sobre la comprobación de hipótesis y, consecuentemente, sobre la adquisición de creencias. No sería necesario recurrir a ningún acto intencional para explicar el autoengaño.

III. EL PROBLEMA SELECTIVO

Recientemente, el modelo intencional de Mele se ha encontrado con una crítica interesante. Siguiendo a Bermúdez [1997, 2000] hay un problema básico que las explicaciones no intencionales del autoengaño no parecen considerar: lo que él llama “el problema selectivo”. Bermúdez señala que el autoengaño es paradigmáticamente selectivo ya que para cualquier caso particular de autoengaño sería necesario explicar porqué el sesgo motivacional tuvo lugar en esa situación particular. Todas las explicaciones se referirían a eventos particulares debido a que el deseo se considera una condición necesaria pero no suficiente para motivar el sesgo cognitivo a favor de la creencia de que p . Esta es la razón por la que “hay muchas hipótesis por las

que mi conjunto de motivaciones señalan un umbral de baja aceptación y alto rechazo y por las que la evidencia que tengo disponible es suficientemente marginal como para hacer posible el autoengaño. Pero me autoengaño cuando llego a creer sólo una pequeña proporción de ellas. ¿Por qué esas y no las otras?” [Bermúdez (2000), p. 318]. Interpreto que Bermúdez se refiere a que el modelo de Mele no es capaz de dar una explicación general del autoengaño. Todas las explicaciones se refieren a casos individuales en momentos particulares. De hecho, esto es lo que Mele afirma cuando lo cité en el último párrafo de la sección anterior.

En relación con la crítica de Bermúdez, Mele no ve ninguna ventaja en explicar el autoengaño como resultado de la intención del que se autoengaña de producir la creencia falsa. Todos nosotros tenemos muchas intenciones, algunas veces llevamos algunas a cabo y en otras ocasiones no. ¿Por qué adquirimos la creencia pertinente en algunos casos y no sucede así en otros? En otras palabras, si la intención juega un papel importante en la explicación del autoengaño, también juega un papel vital en la adquisición de cualquier creencia verdadera. Sin embargo, todos sabemos que la intención no es suficiente para llegar a la creencia. Por tanto, según Mele, “la posición intencionalista enfrenta su propio ‘problema selectivo’” [Mele (2001), p. 66], incluso cuando el autoengaño no es el problema a explicar. Parece como si Mele respondiera a sus críticos con un ataque. Sin embargo, y aunque considera la crítica de Bermúdez seriamente cuando acepta que además del umbral de confianza, los deseos y la evidencia, existen otros factores relevantes para explicar el autoengaño, justifica las limitaciones de su explicación no intencional afirmando que “nadie actualmente puede proporcionar *una explicación adecuada completa de la etiología del autoengaño*. La complejidad del fenómeno sobrepasa nuestro conocimiento” [Mele (2001), p. 66, la cursiva es mía].

Realmente, el umbral de confianza para la aceptación y el rechazo es un mecanismo limitado para explicar el autoengaño. Una explicación más completa tendría que dar cuenta de qué está implicado en estos umbrales. Por ejemplo, mi bajo umbral de aceptación de acciones como violentas (porque soy muy susceptible a la violencia y no necesito de muchas comprobaciones para identificar y aceptar como tales ciertas acciones) no explica por qué tengo un umbral alto de aceptación de las acciones violentas de *mi* hijo, y, como consecuencia, por qué me autoengaño acerca de mi hijo. Comparto con Bermúdez la crítica que hace notar que en el modelo no intencional no queda clara cuál es la conexión entre nuestro deseo de que p suceda y nuestro sesgo motivacional. Sin embargo, considero que la eliminación de la intención como requisito para explicar el autoengaño sí evita paradojas y nos permite prescindir, siguiendo un principio de parsimonia, de un elemento que no esclarece el objeto de estudio. No obstante, necesitamos una explicación más completa acerca de cuál es el origen del sesgo motivacional. En otras pala-

bras, como dije en la sección previa, si el umbral depende del deseo, ¿de qué depende el deseo?

IV. CREENCIAS FALSAS SOBRE UNO MISMO Y AUTOCONOCIMIENTO

IV.1

En cualquiera de los casos cotidianos de autoengaño que podemos imaginar, nuestro deseo no sesga arbitrariamente la evaluación que hacemos de la evidencia y, por ello, llegamos a creencias falsas. El problema selectivo que Bermúdez plantea sólo se puede abordar si consideramos aquello que comparten todos los casos de autoengaño: la relevancia que tiene para el yo mantener una creencia falsa a pesar de la evidencia contraria a la creencia. Trasladar el problema del autoengaño del terreno de la creencia falsa al del papel que juega el yo (utilizaré en lo que sigue el término “yo” para referirme al sujeto que se autoengaña) en ese proceso esclarece el problema selectivo y abre una nueva perspectiva en este asunto.

Holton (2000/1) apunta en esta dirección cuando sitúa al yo como el objeto del engaño y no como sujeto que actúa para engañarse a sí mismo, cambiando de esta forma el lugar que tradicionalmente se le ha dado. Desde su perspectiva, el error acerca del yo es tan esencial en el autoengaño que constituiría una condición necesaria para que se dé éste. La cuestión sería, por tanto, que tenemos falsas creencias sobre nosotros mismos, pero no cualquier clase de creencia falsa constituiría un caso de autoengaño, sino aquella “acerca del tipo de tema que, si estuviéramos en lo correcto, constituiría autoconocimiento” [Holton (2000/1), p. 53].

Al concebir al yo como el objeto del autoengaño se sugiere la relevancia que tiene para el yo considerar como verdaderas creencias no apoyadas por la evidencia. Pero, ¿por qué ciertas creencias acerca de uno mismo son relevantes y otras no? ¿Acaso porque el sentido común nos dice que es más fácil que nos autoengañemos acerca de la fidelidad de nuestras parejas o de nuestra valía profesional que sobre la salud de nuestras plantas o sobre si hemos tomado algún antibiótico durante el último mes? Creo que la respuesta a estas preguntas tendría que explicar qué nos impide en ciertas ocasiones, aquellas en las que sucede el autoengaño, revisar nuestras creencias y ajustar nuestros juicios a la evidencia. Y para ello se requiere un nuevo elemento en esta historia: el papel que la emoción o el afecto juega en el fenómeno del autoengaño.

La emoción juega un papel relevante y está presente en todos nuestros aprendizajes y, por consiguiente, en el procesamiento de cualquier información. Las evidencias no se consideran y, por tanto, no se evalúan de manera neutral. Las situaciones y sus elementos componentes (personas, acciones, sucesos) suelen estar ligadas a valencias positivas o negativas dependiendo

de nuestras experiencias previas, físicas y sociales, con ese tipo de situaciones o de estímulos; es decir, la información está afectivamente marcada [Damasio (1994), Peterson *et al.* (2003), Thagard & Sahdra, (2003)]. A lo largo de nuestra vida hemos ido construyendo conocimiento no sólo acerca del mundo que nos rodea (físico y social), sino también acerca de nosotros mismos. Ambos tipos de conocimiento han sido afectivamente marcados. Tanto la selección como la evaluación de la información en las situaciones cotidianas se realiza siguiendo criterios de coherencia (considerando su valor de verdad y su valencia positiva o negativa), con las creencias previas, especialmente aquellas creencias acerca de uno mismo [Thagard & Sahdra (2003)]. Las informaciones que entran en conflicto en el caso del autoengaño son, por un lado, la que muestra la evidencia y, por otro lado, la que conforma mi autoconocimiento. De aquí que, en vez de considerar la evidencia de manera objetiva y modificar mi autoconocimiento, negamos la evidencia y persistimos en nuestro autoconocimiento. Por eso, el objeto de engaño en el autoengaño no son los datos exteriores, los que recibimos del mundo externo, sino el yo (lo que se cree acerca de uno mismo).

Si el análisis de la etiología del autoengaño sugiere que éste es resultado de un escaso o falso autoconocimiento, si aumentáramos nuestro conocimiento, disminuiría el autoengaño. Otra cosa es la deseabilidad o no del fenómeno del autoengaño. Sobre este punto no hay acuerdo ni en la literatura filosófica ni psicológica, pero podríamos considerar que, al menos desde el punto de vista de la epistemología, la aspiración a alcanzar el mayor número posible de creencias verdaderas, no hace de los estados de autoengaño nuestro objeto de deseo. Mis reflexiones finales abren la puerta a consideraciones normativas en torno a este tema.

IV.2

Si el conflicto que se produce en el autoengaño proviene de la contradicción entre lo que muestra la evidencia y lo que sé de mí, mi autoconocimiento, entonces ambos aspectos influirán en la eliminación del autoengaño.

Por un lado, a pesar de que la carga afectiva de la evidencia ignorada en el autoengaño impide considerarla, desde el punto de vista del sujeto que se autoengaña, como un elemento objetivo de la situación, la evidencia juega en contra de la pervivencia del autoengaño: cuanta más evidencia sobre la que autoengañarse reciba el sujeto que se autoengaña, más difícil será mantener el autoengaño. El grado de esta dificultad dependerá de la forma en que esa evidencia se presente al sujeto. Si son datos que sólo de manera inferencial contradicen la creencia objeto del autoengaño su peso será menor que si son datos que no requieren de inferencia alguna, datos a los que podemos acceder a través de la observación directa y en el momento presente. Por muy subjetivamente interpretable que sea la información, alguna lo es más que otra.

Por otro lado, además de la evidencia, el otro elemento a tener en cuenta a la hora de intervenir para reducir el autoengaño ha de ser nuestro autoconocimiento. Cuanto mayor sea nuestro autoconocimiento, menos susceptibles seremos al autoengaño. El autoconocimiento implica que los sujetos conozcan como son sus respuestas emocionales y como pueden modificar sus emociones a medida que surgen [Feldman Barrett & Gross (2001)].

En conclusión, aunque en el nivel descriptivo-explicativo he rechazado los modelos intencionales de autoengaño (Davidson) porque originan paradojas de difícil solución, la alternativa de Mele tampoco es satisfactoria, ya que no es el deseo de que p el que me lleva a evaluar erróneamente la información que indica la verdad de $\neg p$, sino que es el desconocimiento de la carga afectiva de una información, que se sitúa en el yo, lo que me lleva a ignorarla y a autoengañarme. Si el yo es el objeto de engaño, el autoconocimiento podría reducir nuestro nivel de autoengaño. En este punto, recuperamos la intencionalidad (nuestro objeto de discusión inicial), ligada a la autoconciencia. En este tema, como en otros muchos, el precepto delfico, “conócete a ti mismo”, vuelve a ser pertinente y actual.

*Departamento de Historia y Filosofía de la Ciencia, la Educación y el Lenguaje
Universidad de la Laguna
Campus de Guajara, E-38200, La Laguna
E-mail: roherbo@ull.es*

REFERENCIAS BIBLIOGRÁFICAS

- BARNES, A. (1997), *Seeing through self-deception*, Cambridge, Cambridge University Press.
- BERMÚDEZ, J.L. (1997), “Defending Intentionalist Accounts of Self-Deception”, en *Behavioral and Brain Sciences* 20, pp. 107-108.
- (2000), “Self-deception, intentions and contradictory beliefs”, en *Analysis*, 60, pp. 309-319.
- CHISHOLM, R. M. (1966), “Freedom and Action”, en Lehrer, K. (ed.) *Freedom and determinism*, New York, Random House.
- DAMASIO, A. (1994), *Descartes' Error. Emotion, Reason and the Human Brain*, Nueva York, Grosset/Putnam.
- DAVIDSON, D. (1969), “How is weakness of the will possible?”, en Davidson, D., *Essays on Actions and Events*, D. Davidson, 1980, Oxford, Oxford University Press.
- (1980), *Essays on Actions and Events*, Oxford, Oxford University Press.
- (1982), “Paradoxes of Irrationality”, en Hopkins, J. y Wollheim, R. (eds.) *Philosophical Essays on Freud*, Cambridge, Cambridge University Press.
- (1986), “Deception and Division”, en Elster, J. (eds.), *The Multiple Self*, Cambridge, Cambridge University Press.

- (1998), “Who is fooled?”, en Dupuy, J.P. (ed.), *Self-Deception and Paradoxes of Rationality*, Stanford, CSLI Publication.
- FELDMAN BARRET, L. y GROSS, J. (2001), “Emotional Intelligence. A Process Model of Emotion, Representation and Regulation”, en Mayne, T. & Bonanno, G., *Emotions. Current Issues and Future Directions*, Nueva York, The Guildford Press.
- FRIEDRICH, J. (1993), “Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena”, en *Psychological Review*, 100, pp. 298-319.
- HOLTON, R. (200/1), “What is the role of the self in self-deception?”, en *Proceedings of the Aristotelian Society*, 101, pp. 53-69.
- MELE, A. R. (1987), *Irrationality. An Essay on Akrasia, Self-Deception, and Self-Control*, Oxford, Oxford University Press.
- (1998), “Two Paradoxes of Self-Deception”, en Dupuy, J.P. (ed.), *Self-Deception and Paradoxes of Rationality*, Stanford, CSLI Publication.
- (2001), *Self-Deception Unmasked*, Princeton, Princeton University Press.
- PETERSON, J.B., et al. (2003), “Self-Deception and Failure to Modulate Responses Despite Accruing evidence of Error”, *Journal of Research in Personality*, 37, pp. 205-223.
- SEARLE, J. (1983), *Intentionality. An Essay in the Philosophy of Mind*, Cambridge, Cambridge University Press.
- THAGARD, P. y SAHDRA, B. (2003), “Self-Deception and Emotional Coherence”, en *Mind and Machines*, 13, pp. 213-231.
- TROPE, Y. y A. LIBERMAN (1996), “Social Hypothesis Testing: Cognitive and Motivational Mechanism”, en E. Higgins and A. Kruglanski, (eds.), *Social Psychology: Handbook of Basic Principles*, Nueva York, Guilford Press.