MODOS DE AUTOENGAÑO Y DE RAZONAMIENTO: TEORÍAS DE PROCESO DUAL*

SALMA SAAB
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México
salma@unam.mx

Resumen

En este artículo me ocupo de la cuestión de cómo en las teorías de proceso dual se puede dar cuenta del autoengaño y su conexión con la racionalidad. Presento las versiones intencionalista y no intencionalista del autoengaño y muestro cómo el debate entre ellas puede dirimirse de manera más completa y satisfactoria en el marco de una teoría dual. En éste suelen aceptarse dos sistemas de razonamiento, uno heurístico (S_1) y otro analítico (S_2), que compiten por el control de nuestras inferencias y acciones, pero a veces interactúan y colaboran entre sí. Se defiende que si predomina la respuesta de S_1, se puede ver el patrón del autoengaño como una forma de razonamiento heurístico y no únicamente como un vínculo causal. Se sugiere que las evaluaciones en cuanto a la racionalidad del proceso del autoengaño, dependerá del modo en que intervenga en el patrón de razonamiento y del sistema desde el cual se lleve a cabo.

PALABRAS CLAVE: Autoengaño; Racionalidad; Razonamiento; Intencionalidad; Teorías de proceso dual.

Abstract

In this paper I discuss the phenomenon of self-deception and its connection with the notion of rationality linked to the dual process theories. I present the intentionalist and non-intentionalist accounts of self-deception and aim to show how the debate between them can be resolved in a more comprehensive and satisfactory manner, if it is placed in the frame of the dual process theories. The dual model usually accepts two kinds of reasoning processes, heuristic and analytic, referred to two different systems, S_1 and S_2. These processes compete for control of our inferences and actions, but sometimes they interact and collaborate. It is suggested that in a dual model, the evaluations in terms of the rationality of the process will depend on the way in which self-deception participates in the reasoning process and on the system from which the evaluation takes place.

KEY WORDS: Self-deception; Rationality; Reasoning; Intentionality; Dual process theories.

* Agradezco a Axel Barceló, Raúl Quesada y a los dictaminadores anónimos de la revista por sus valiosos comentarios y a Diana Pérez por sus útiles sugerencias a una versión anterior de este trabajo. También agradezco a Ana Segovia por sus correcciones en la redacción del texto.

Introducción

La mayoría de los autores que se ocupan del autoengaño coinciden en entender el autoengaño como un fenómeno en el que irrumpen, de algún modo, los deseos o emociones de la persona en el proceso de formación de sus creencias, de tal manera que la creencia a la que se llega es un tipo de creencia falsa. Así, se supone que en el autoengaño se trastoca o altera el proceso de formación de creencias, violando con ello los estándares apropiados para sustentarlas. Al admitirse que en el autoengaño está involucrado un proceso de formación de creencias, se abre la puerta a la pregunta de si es adecuado. Es frecuente que se juzgue –tanto en un contexto cotidiano como filosófico– que en el proceso inferencial en el que intervienen elementos motivacionales se violan los estándares apropiados para sustentar creencias, por tanto, que se evalúe como irracional. Se considere como irracional la creencia a la que se llega o el proceso que la engendra, en ambos casos se juzga que alberga una suerte de paradoja. Sin embargo, algunos están en desacuerdo con la valoración del autoengaño como irracional.

En este trabajo me centro en los procesos y en las versiones más usuales del autoengaño que ilustran las valoraciones antagónicas en cuanto a su racionalidad: la intencionalista y la no intencionalista. Intento mostrar que ninguna de ellas captura satisfactoriamente el fenómeno del autoengaño y sugiero que podemos dar una mejor caracterización del autoengaño en el marco de una teoría dual de sistemas. El resultado de insertar su discusión en esta permite que integremos rasgos de ambas teorías ya que el mismo ámbito de esta teoría asume algún tipo de división –al igual que lo hacen los intencionalistas–, pero, al estar acompañada de dos formas de razonamiento, da cabida a que cambie la valoración del autoengaño en cuanto a su racionalidad según se explique bajo una u otra de dichas formas. A la vez, del mismo modo en el que lo hacen los no intencionalistas, se confiere al autoengaño una función psicológica que aporta beneficios a la especie humana y al individuo.

Procederé de la siguiente manera. En la primera sección presento las versiones intencionalista y no intencionalista del autoengaño y me enfoco únicamente en la discusión de dos de sus rasgos centrales: la intencionalidad y la racionalidad. En la segunda sección distingo diferentes nociones de racionalidad y la manera en la que la racionalidad se conecta con el razonamiento. En la tercera me refiero a la teoría dual de razonamiento y la manera en que en ella se explican y reinterpretan los resultados de variadas pruebas experimentales de razonamiento a las

que se sometieron muchos individuos y en las que la mayoría respondió sistemáticamente de manera equivocada, pretendiéndose entonces invalidar la hipótesis de que los humanos somos irracionales, conclusión que los investigadores extrajeron de esos resultados. Sin embargo, con la propuesta de la teoría dual de razonamiento, sus defensores intentan acotar el juicio de la irracionalidad humana. En la cuarta y última sección parto de la propuesta no intencionalista de Johnston y recojo su idea de que el autoengaño tiene una función en nuestra vida psicológica, que es sub-intencional, pero difiero de considerarla un mero mecanismo causal o disposicional. Señalo que cuando se sitúa la discusión del autoengaño en el marco de una teoría dual, logramos verlo como un proceso que puede involucrar un desarrollo de razonamiento que, según el sistema de pensamiento del que provenga, podría evaluarse como racional o irracional.

I

Como existen otros fenómenos diferentes que pueden compartir con el autoengaño la descripción de que la creencia que se obtiene se basa en motivaciones del sujeto –como por ejemplo el pensamiento esperanzado o desiderativo (*wishful thinking*)–, muchos autores consideran que tendríamos que ofrecer una manera de diferenciarlos. La forma más común de especificar el autoengaño es la siguiente: la adquisición y retención de una creencia (o al menos admitirla) frente a evidencias en su contra, motivada por deseos, emociones o intereses del sujeto que favorecen su adquisición y retención.

Algunos ejemplos de autoengaño estructuralmente pertinentes en esta discusión son los siguientes: 1) Una persona que tiene fuertes evidencias de que su cónyuge le es infiel, junto con el deseo de querer paliar el dolor o la ansiedad que le produce que la infidelidad tuviera sustento, cree que le es fiel. 2) Un sentido contrario al anterior, sería el de una persona celosa que, frente a escasa evidencia de infidelidad, aunada a su temor y angustia de que la hubiera, tenga continuas sospechas que la llevan a creer la supuesta infidelidad.[1] 3) Una persona que no acepta que tiene una enfermedad incurable, aunque se la hayan diagnosticado varios médicos. 4) Alguien que considera a un ser querido

[1] Estos casos se conocen en la literatura como casos de autoengaño retorcido (*twisted*) ya que en la persona se produce un estado que le causa mayor dolor, contrariamente al efecto que en general se busca con el autoengaño, paliar el dolor o la ansiedad que genera el que fuera verdadera (véase Mele 1999, 2000).

inocente de haber cometido algún ilícito, a pesar de las múltiples evidencias y de que los jueces lo hayan encontrado culpable, ya que le produciría un gran dolor y sufrimiento admitirlo.

Algunos investigadores experimentales defendieron la hipótesis de la presencia de la irracionalidad humana aún en casos en los que no intervenían elementos motivacionales. Así, en las décadas de los años sesenta y setenta del siglo pasado –en las que se dio la revolución cognitiva en el campo de la psicología–, estos investigadores diseñaron y aplicaron, incluso a sujetos con instrucción universitaria, toda una batería de experimentos que involucraban pruebas de razonamiento bastante simples. Los resultados que obtuvieron evidenciaron que la mayoría de los sujetos analizados tendían sistemáticamente a tomar decisiones equivocadas o irracionales. Con base en el sesgo cognitivo que arrojaron las pruebas, los investigadores extrajeron un resultado muy sorprendente y desolador: la hipótesis de que las personas son sistemáticamente irracionales. Los experimentos se conocen en la literatura como "paradojas de la racionalidad". (Para darse una idea de los tipos de experimentos que se llevaron a cabo y los resultados extremos a los que llegaron, véase Kahneman y Tversky 1982; Gilovich, Griffin y Kahneman 2002, y Piatelli-Palmarini 2005).

A raíz de estos estudios, entre otros, creció el interés, tanto de filósofos como de psicólogos, de repensar la noción de racionalidad, al igual que los conceptos relacionados con ella, como por ejemplo el de razonamiento y justificación. El impulso mayor se dio en los estudios que llevaron a cabo Evans y Over (1996), circunscritos al estudio del razonamiento y el juicio, y que cristalizaron en la propuesta de la teoría dual de la racionalidad.[2] Con esta teoría pretendieron resolver las paradojas de la racionalidad y contrarrestar aquel resultado pesimista. En síntesis, para estos autores, el tipo de conflicto que se evidencia en las respuestas dadas a tales tareas de razonamiento se interpreta y explica no porque razonemos incorrectamente sino porque existen diferentes formas de procesar la información (Evans y Over 1996). Como elaboraré más adelante, ellos propusieron la existencia de dos formas de razonamiento, una *instrumental* o *heurística* (r_1) y otra *lógica* o *abstracta* (r_2) y también sugirieron de qué modo pueden relacionarse.

En este trabajo persigo retomar el debate entre los intencionalistas

[2] En otros campos de estudio dentro de la psicología surgieron, de manera independiente, otras propuestas de teorías duales. Para tener una idea de las diferentes versiones de la hipótesis dualista véase el recorrido histórico que presentan Evans y Frankish (2009).

y los no intencionalistas en torno al autoengaño y discutirlo a la luz de los debates que se llevan a cabo en la ciencia cognitiva experimental. En particular, en el marco de las teorías de sistema dual –las más aceptadas en la actualidad–, presento la manera en que enfrentan el problema de la racionalidad e irracionalidad humana y planteo una alternativa en la que se puede insertar el entendimiento del fenómeno del autoengaño en este tipo de teoría.

Los diversos intentos de analizar el fenómeno del autoengaño se han topado con la dificultad de sortear las diferentes paradojas que parece entrañar su descripción. La mayoría de los diferentes propósitos de eliminar ese aire de paradoja se pueden integrar en dos grupos: el intencionalista y el no intencionalista. El autoengaño se puede atribuir al estado o al proceso y suelen distinguirse dos tipos de irracionalidad. Mele considera que la irracionalidad asociada al estado genera una paradoja que denomina "estática", mientras que la irracionalidad asociada al proceso produce una paradoja que denomina "dinámica" o "estratégica".[3] Sin embargo, en este trabajo, me ocupo únicamente del caso del proceso.

Las diferencias entre estas dos posiciones y sus variantes son múltiples, pero aquí sólo me detendré en sus posturas en relación con dos cuestiones: la intencionalidad y la racionalidad. En primer término, cabe señalar que difieren en cuanto a si el autoengaño es un asunto de cognición o de volición. Si se responde que se trata de un asunto de cognición y se asume que el autoengaño es intencionalmente inducido por el sujeto –como sugieren los intencionalistas–, se recurre a la fragmentación del sujeto para evitar las paradojas; no obstante, de cualquier forma, se llega a la conclusión de que el autoengaño es irracional. Si se considera como una cuestión de volición, como suelen suponer los no intencionalistas, donde el sujeto de algún modo no es consciente del engaño (ya que se configura en un nivel preconsciente o no consciente), entonces se abre la posibilidad de que no se considere irracional o, acaso, no racional. Elaboraré este punto en la última sección del trabajo.

Para los intencionalistas, lo central del autoengaño es que la persona *intencional* o deliberadamente se induce a sí misma una creencia que ella cree (o sabe) que es falsa. El autoengaño constituye una falla cognoscitiva: se asume que la forma apropiada de generar creencias, o de inferirlas, debe hacerse con apego a los hechos y con base en las evidencias disponibles. Se considera que hay una estrecha conexión entre la adecuada formación de creencias y la búsqueda de la verdad. De este

[3] Mele (2001, 1987).

modo, la falla que el sujeto comete al hacer que intervengan sus deseos o emociones se explica en términos de que éste *deliberadamente* sesga la evidencia en favor o en contra de su creencia, según su conveniencia, corriendo el riesgo de producir creencias falsas o irracionales.

Para algunos intencionalistas, en el autoengaño el sujeto tiene creencias contradictorias (Davidson 1992). Así, recurren al modelo de fragmentación o compartimentación de la mente, para explicar que un sujeto pueda tener creencias inconsistentes pero, de algún modo, separadas entre sí.[4] Sin embargo, para otros, no es necesario atribuirle al sujeto creencias contradictorias, aunque conservan la idea de que la persona tiene evidencias que apoyarían la creencia contraria a la que explícitamente sostiene.

Además, para esta corriente, el sujeto es considerado epistémicamente responsable e incluso moralmente responsable de sus creencias, al igual que de sus consecuencias. El sujeto comete una falta moral en la medida en que doblega sus creencias en la dirección de lo que desea creer.

Para los no intencionalistas, al no ser el autoengaño ni intencional ni consciente, consideran que puede cumplir otra función en la economía psicológica de la persona, muy distinta a la de generar verdades. Se podría no exigir asentimiento para todo lo que consideramos verdadero. Pensemos en casos en los que lo que tomamos como verdadero se vincula con "pensar positivamente". Sería una forma de "desear creer" que tiene la función de incrementar las probabilidades de que realmente se cumpla la creencia deseada y no la opción contraria que tanto se teme. En este pensar positivo se recurre a mecanismos habituales en los que se presta atención a, o se focaliza en, cierto tipo de información y no a otra, esto es, se selecciona la información. Si se da el caso de que se reconozca información adversa a lo que el sujeto quiere creer, éste la ignora o desestima. Sostienen también que, en la mayoría de los casos, el autoengaño obedece a pautas de conducta regulares y no accidentales en

[4] Davidson sostiene que una vez que se admite la *irracionalidad del estado* en el que se encuentra el autoengañado, ya que contiene creencias inconsistentes, surge la pregunta por el punto de la secuencia que hace posible que se llegue al estado irracional. Davidson responde que el paso irracional que lo hace posible se da al "trazar la frontera que separa a las creencias inconsistentes. Cuando el autoengaño consiste en debilidad de la justificación autoinducida, lo que debe clausurarse del resto de la mente es el requisito de los elementos de juicios totales. Lo que causa su exilio o aislamiento temporal es, claro está, el deseo de evitar aceptar lo que aconseja el requisito. Pero esto no puede ser una *razón* para descuidar el requisito. Nada puede verse como una buena razón para *no* razonar de acuerdo con nuestros mejores patrones de racionalidad" (1992, p. 100 [versión en español]).

las que las ansiedades, miedos y otros sentimientos generan creencias (Johnston 1988; Mele 2000 y Barnes 1997). Algunos, como Johnston, no concluyen que estas se generen de manera anómala o patológica, sino que se debe a la activación de una disposición mental natural.[5] Constituiría un mecanismo que tiene un propósito, pero de tipo sub-intencional.

Podríamos sintetizar las diferencias de las dos posturas con respecto al autoengaño de la siguiente manera, aunque no todos los rasgos que menciono sean aceptados unánimemente por sus defensores:

A) Intencionalistas

i) Fenómeno cognitivo que se explica en términos de creencias. Por lo general, apela a un modelo jerárquico, en el cual las creencias de primer orden pueden ser inconscientes, mientras que las de segundo orden, conscientes.
ii) El autoengaño es intencional. El sujeto tiene la intención de creer *p*.
iii) Existe una liga entre creencia y verdad. Sin embargo, la intención de creer *p* no se basa en evidencia alguna que apoye la verdad de *p*.
iv) El sujeto admite creencias contradictorias, la que intenta negar (que mantiene aislada o segregada) y la que tiene la intención de creer. Para ello, se requiere alguna forma de fragmentación de la mente. Supone de alguna manera mentirse, como en el caso del engaño a los otros, y es a la vez inducido y ocultado por el propio sujeto.
v) Cae dentro del dominio de lo racional: apela a principios normativos, de modo que, al violarlos, la creencia inducida se juzga como irracional.

B) No intencionalistas

i) Fenómeno volitivo que tiene una función que no es necesariamente cognitiva. Es una forma de "asumir" una creencia falsa en la que, de manera motivacionalmente sesgada, irrumpen deseos e intereses del sujeto.

[5] Johnston llama "tropismo mental" al mecanismo mental que tiene una conexión característica entre deseo y creencia, que no es accidental y que no es racional y que constituye la base causal que da soporte tanto a conexiones racionales como irracionales (1988, p. 67).

ii) Obedece a mecanismos habituales en los que se presta atención a, o se focaliza en, cierto tipo de información y no a otra.
iii) Si se admite un tipo de conducta que tiene un propósito, podría considerarse como no intencional, o quizás sub-intencional, que puede tener una función psicológica particular, por ejemplo la de aminorar miedos y ansiedades logrando servir para incrementar la obtención de un fin deseado. En este sentido, adoptar la creencia sería una forma de *pensar positivo.*
iv) Permite que se puedan generar creencias en las que participan las motivaciones del sujeto de manera *adecuada o confiable*, y no de manera accidental, según las metas del sujeto.
v) No es necesariamente irracional.

En apoyo a las teorías de sistema dual, además de la evidencia que proviene de los experimentos que se llevaron a cabo en las tareas de razonamiento y de toma de decisiones, también se ha utilizado evidencia de otras áreas de estudio. Por ejemplo, las correspondientes a la psicología del desarrollo (pospiagetistas), a la psicología cognitiva, social, evolutiva y de las neurociencias. Los análisis evolutivos permiten sustentar la hipótesis que incluye al autoengaño entre las capacidades mentales que adquirimos histórica y evolutivamente, las cuales coadyuvaron a nuestra supervivencia como especie y generaron tipos de mecanismos muy primitivos, desarrollados por nuestros ancestros frente a su entorno natural y social, y que todavía conservamos. Sugieren asimismo que la estrategia del autoengaño surge tras haber desarrollado la capacidad de engañar a los otros y, consecuentemente, de incrementar las capacidades para detectarlo por parte de las víctimas. Consideran que el autoengaño aparece como una estrategia para hacer más efectivo el engaño a los otros. (véase Krebs y Dawkins 1984; Krebs y Denton 1997; Hauser 1996; Trivers 1971; Krebs, Denton y Higgins 1988, y Lockard y Paulus 1988). Estas hipótesis, de ser correctas, reforzarían que se matizara la catalogación del autoengaño como irracional

II

Antes de pasar a la cuestión de la manera en que el autoengaño se puede explicar en el contexto de una teoría de sistema dual y el tipo de desviación o sesgo que involucra, me detendré brevemente en el concepto de racionalidad. En este apartado me referiré a los diferentes casos a los que se aplica el criterio de racionalidad, cómo se conectan entre sí y los diversos parámetros que se han utilizado para evaluarlos.

Considero, como muchos, que –de manera afín a nuestro entendimiento cotidiano y pre-teórico de racionalidad– el que las personas se juzguen por lo general racionales, no excluye que no lo sean en dominios más específicos, en los que se requieren capacidades que se miden según diferentes parámetros: como el del cálculo de probabilidades, las reglas con condicionales o las que guían el razonamiento matemático. Por otra parte, también sugiero, aunque no lo desarrollo en este trabajo, que los elementos motivacionales, lejos de constituir elementos disruptivos de un buen razonamiento, pueden ser esenciales a él.

Una dificultad central con el concepto de racionalidad es que se trata de una noción muy general y no definida, además de que el ámbito de su aplicación puede ser muy diverso. Adicionalmente, tampoco parece haberse constreñido adecuadamente, como lo expresan Botterill y Carruthers, pues los constreñimientos filosóficos que se han ofrecido de la irracionalidad son muy endebles (1999, p. 106).[6] No obstante, estas debilidades en torno a las nociones de racionalidad e irracionalidad no han impedido sus profusos usos.

Botterill y Carruthers distinguen varias nociones de racionalidad epistémica e indican las diferentes maneras en que se pueden relacionar. En primer término, se encuentra la *racionalidad del sujeto* que, como su nombre lo indica, se aplica a la persona íntegra. En segundo término, la *racionalidad del estado mental*, aplicada a la creencia misma o a la actitud epistémica de que se trate. Y, finalmente, la *racionalidad del proceso*, referente al proceso de formación de la creencia o actitud epistémica. A su vez, el estado mental se podría subdividir en un estado racional de *tipo* o de *instancia*. Además de estas nociones de racionalidad, se agrega también la racionalidad del propósito[7]. Sin embargo, de estas diferentes formas de racionalidad –del sujeto, del estado mental y del proceso– la fundamental parece ser la noción de *proceso* racional de formación de creencias para, en función de esta, poder explicar tanto la racionalidad del sujeto mismo como la racionalidad de los estados mentales (*ibid.,* pp. 106-107).

Un proceso de formación de creencias es racional cuando se utiliza un procedimiento que validamos o reconocemos como legítimo, en función de los objetivos o metas que se buscan. Así entendida, la racionalidad nos

[6] Botterill y Carruthers (1999, p. 105).

[7] Se puede hablar de la racionalidad del *propósito* de la siguiente manera: una racionalidad en la que actuamos de manera que nos sirve para alcanzar ciertos fines, que se puede entender como una racionalidad instrumental. Sin embargo, para lograrlos, debemos actuar de manera apropiada (véase Evans 2010, p. 187).

acerca a una postura consecuencialista del razonamiento. En cuanto a la persona, se puede decir que esta es racional, en cuanto a sus creencias, si para sustentar la mayoría de ellas emplea *procesos* de formación de creencias que por lo general se consideran racionales; por ejemplo, el empleo del razonamiento lógico, el razonamiento práctico o la inferencia a la mejor explicación. Si la racionalidad se aplica a los estados mentales cognitivos, como las creencias, decir que una creencia es racional depende, en primer lugar, del tipo de creencia o de una instancia particular. Si se refiere a una creencia-tipo, por ejemplo, que los bebés prematuros necesitan cuidados especiales en incubadoras, esta podría formarse a partir de un proceso racional. Pero, podría darse el caso de que una instancia particular de esa creencia surgiera de un proceso irracional, por ejemplo, si es inducida por un proceso posthipnótico. Así, sería irracional que una persona sostuviera esa instancia de creencia, por no haberla originado de manera racional, esto es, porque no ha utilizado un procedimiento para generar creencias que reconozcamos como legítimo.

Así, si se da primacía a los procesos, muchos autores considerarán natural conectar la racionalidad con el razonamiento. Y si se acepta una teoría dual, la racionalidad dual –una heurística y otra analítica o lógica (o deóntica)– se reflejará, hasta cierto punto, en una teoría dual del razonamiento (Evans y Over 1996, p. 141).

Un supuesto muy difundido en la epistemología consiste en tomar a la lógica y sus verdades como único criterio del cual se deriva la corrección de los patrones de razonamiento, así como los de justificación. Goldman, por ejemplo, cuestiona esta derivación; de los principios formales no pueden derivarse los principios normativos que rigen al razonamiento (Goldman 1986, cap. 5). Una razón que esgrime es que si lo que es propio de las reglas de razonamiento es el tránsito entre estados psicológicos, y los principios lógicos no se ocupan de los estados psicológicos y de sus relaciones, entonces la lógica no puede establecer nada respecto del razonamiento. Otros, como Evans y Over (1996), únicamente señalan las limitaciones de los estándares formales para evaluar un buen razonamiento o una buena toma de decisión (*ibid.*, p. 142). Volveré a esta cuestión en la tercera sección.

Con respecto a la noción de racionalidad que suele emplearse en discusiones filosóficas, se la asocia a la capacidad, específicamente humana, para pensar y razonar o hacer inferencias. Contrariamente a esta visión de los filósofos, en el campo de la psicología cognitiva y de estudios comparativos con otras especies animales, se proponen modelos en los que se puede dotar a otras especies de capacidades de razonamiento, aunque más limitadas que las de los humanos. En la teoría dual, las

atribuciones de capacidades de razonamiento a otras especies serían semejantes a las atribuidas a los humanos, las cuales estarían asociadas al sistema S_1.

En cuanto a la racionalidad del proceso, muchos filósofos parten de un modelo de racionalidad idealizado, en el cual se establecen los principios generales normativos de cómo *debemos* pensar y, con base en ellos, se juzga nuestra racionalidad humana. Davidson, por ejemplo, parece suscribir ese paradigma de racionalidad. Davidson utiliza el concepto de racionalidad como inherente a su proyecto de intepretación de la conducta de los otros. Nuestra habilidad para entender a los demás y de verlos como agentes presupone un principio que Davidson denomina principio de caridad o de humanidad. Este principio supone que el sujeto de interpretación es racional, es decir que la mayoría de sus creencias son verdaderas, antes de poder atribuirle nuestros propios conceptos psicológicos ordinarios (creencias, deseos e intenciones y otras actitudes proposicionales). La racionalidad como prerrequisito para poder interpretar al otro que utiliza Davidson está relacionada con el concepto de racionalidad aplicado al sujeto que distinguen Botterill y Carruthers. En este trabajo no me ocupo de este concepto de racionalidad y tampoco del problema del agente[8], que si bien son problemas importantes y tienen implicaciones para lo que aquí defiendo, rebasa los límites de los objetivos que aquí persigo. Algunos también imputaron a los psicólogos cognitivos el uso de la noción de racionalidad idealizada y consideraron que, por ello, llegaron a la conclusión de la irracionalidad humana.

Muchos sugieren que el estándar de racionalidad frente al cual se debe medir el desempeño humano, en lugar de ser el estándar idealizado, debe ser relativo a las habilidades de los humanos (Cherniak 1986; Botterill y Carruthers 1999), pero también a nuestras necesidades en tanto que pueden existir presiones de tiempo o propósitos específicos en los que se exige una respuesta rápida. Los defensores de la teoría dual se suman a esta forma de entender la racionalidad.

Esta noción de racionalidad relativa a nuestras necesidades y capacidades, en tanto seres finitos en búsqueda de la verdad, está ligada a la noción de razonamiento relativo, esto es, como algo que hacemos más o menos bien, lo cual permite que nos alejemos de la forma de razonamiento inferencialmente válido de la lógica. Con base en esta

[8] Sturm (2007), por ejemplo, considera que el rol del agente es indispensable para el entendimiento del autoengaño y critica a los no intencionalistas de suprimir al agente como autor del autoengaño y de ese modo intentar evitar la consideración del autoengaño como resultado de un proceso de deliberación práctica.

distinción, los defensores de la teoría dual pueden ofrecer constreñimientos cognitivos más adecuados para una racionalidad r_1, en donde el razonamiento puede ser defectuoso o sesgado, cuando la persona tiene limitaciones en sus capacidades cognitivas o tiene pocos recursos disponibles (véase Evans y Over 1996, p. 142). En cambio, la racionalidad r_2 es defectuosa cuando se desvía de los principios normativos formales.

Esta manera relativa de entender la racionalidad va de acuerdo con nuestro entendimiento pre-teórico de racionalidad, el cual sugiere que las personas pueden considerarse en general racionales y no serlo en dominios más específicos, por ejemplo, el razonamiento con condicionales, con probabilidades o inferencias a la mejor explicación. El desempeño de una misma persona puede variar en cada uno de estos casos. Debemos distinguir entre cómo *debemos* razonar y cómo, de acuerdo con ciertos parámetros, *de hecho* razonamos (Botterill y Carruthers 1999, pp. 105-106). Puede ser que los dos procesos mentales coincidan en el mismo resultado, pero también que difieran; en caso de diferir, parece que debe haber un límite en cuanto a qué tanto se puede tolerar que difieran o que pueda propiamente seguir denominándose "razonamiento".

En la actualidad se observa otro giro importante en relación con la racionalidad, que tiene que ver con el papel que en ella juegan los elementos motivacionales, como los sentimientos, los deseos y las emociones. En la versión estándar de la racionalidad, los elementos motivacionales son disruptivos –entorpecen y desvían– del buen funcionamiento de la razón. En la revisión actual, que se da también en otros campos de estudio de la mente, se destaca la estrecha conexión entre la racionalidad y las emociones. Con este giro se da lugar a un rango de juicios emotivo-cognitivos involucrados en la explicación de ciertos fenómenos psicológicos, entre los que estaría el autoengaño (Epstein 1994; Hassin *et-al.* 2005).[9]

En otra sugerente propuesta en esa misma dirección, pero dentro de un modelo computacional, se modela el autoengaño como el resultado de lo que se denomina "coherencia emocional", dirigido a aproximarse o evitar objetivos subjetivos (Sahdra y Thagard 2003). Se considera que los efectos de las emociones podrían estar mediados por el pensamiento racional, del mismo modo que, a la inversa, los efectos del pensamiento racional podrían estar mediados por emociones. Pero, en este trabajo no me detendré en el desarrollo y discusión de estos giros.

En cuanto a la distinción entre la predicación de la racionalidad cuando esta se aplica al estado o al proceso, también se plantea –como ya mencioné con anterioridad– en la discusión del autoengaño. Pero, a

[9] Véase Evans y Frankish (2009).

diferencia de lo que sucede cuando se definen en el caso de la racionalidad, cuando se trata del autoengaño no se aborda la cuestión de cómo se conectan. Así sucede, por ejemplo, en los casos que mencioné con anterioridad (de Davidson y Mele). Por ejemplo, en el de Davidson, a la irracionalidad del estado sólo se le suma la irracionalidad del proceso (véase nota 4 de este trabajo). Y en el caso de Mele, se distinguen dos tipos de irracionalidad: la asociada al estado y la asociada al proceso, pero no se dice nada respecto de su relación.

Pasemos, entonces, a la solución propuesta en las teorías de proceso dual y cómo en relación con estas replanteo la discusión del autoengaño.

III

La teoría dual de la cognición se ha propuesto en diferentes campos de estudio. Ya mencioné antes la psicología del desarrollo y la psicología evolutiva, y se pueden agregar la psicología del aprendizaje, la memoria y la psicología social.[10] Sin embargo, se la ha defendido de muy diversas maneras, en el sentido de que cada uno de los sistemas teóricos no siempre contiene las mismas características centrales. En esta sección, presento los rasgos de la teoría dual tal como se defiende en relación con el razonamiento y la toma de decisiones. Una de las teorías más representativas en estos últimos campos es la que proponen Evans y Over (1996), aunque al paso de los años ha sufrido modificaciones importantes.[11] En este trabajo me apoyo primordialmente en la teoría dual en la forma en que la defiende Evans en sus escritos más recientes (2008, 2009 y 2010).

En un inicio, como ya indiqué, una de las razones principales que llevaron a Evans y Over al desarrollo de la teoría de proceso dual fue explicar el conflicto de respuestas que se observa en los experimentos de laboratorio en los que se evaluó el desempeño de los sujetos en varias tareas de razonamiento y de toma de decisiones. En algunos experimentos se puso a prueba su desempeño en el razonamiento deductivo, proporcionándoles ejemplos de silogismos, unos válidos y otros inválidos, con contenido variado: a veces incluían una conclusión creíble y otras, una

[10] Un ejemplo, en el campo de la neurología social cognitiva, es Lieberman (2003 y 2009), con su distinción entre el sistema *reflejo* (S_1), denominado Sistema-X, y el sistema *reflexivo* (S_2), denominado Sistema-C, y cuyo funcionamiento nos remite a diferentes regiones del cerebro.

[11] Otras propuestas duales son las de Sloman (1996) y Stanovich y West (1998).

no creíble. Como resultado obtuvieron que el contenido modificaba o sesgaba la evaluación del razonamiento en una dirección diferente a la que se tiene cuando sólo se juzga por la forma o estructura del silogismo y que los sujetos tendían a aceptar más frecuentemente como válidos los razonamientos con conclusiones más creíbles que los que contenían las conclusiones menos creíbles. También aceptaban con más frecuencia las conclusiones válidas que las inválidas, además de que en los argumentos inválidos se mostró una mayor desviación (Evans 2003, p. 455).

Evans y Over estudiaron estos supuestos sesgos y errores cognitivos. La mayoría de los autores de la época consideraron que permitir cualquier influencia de este tipo era normativamente incorrecto. Si en los experimentos se les indica explícitamente a los participantes que tienen que evaluar la validez de los argumentos y juzgar si lo que dice la conclusión es falso o es verdadero, se producen contestaciones divergentes frente a un mismo patrón deductivo, produciéndose fallas sistemáticas a pesar de que la cuestión del contenido debería ser irrelevante para la respuesta.

Evans y Over consideran equivocado que el desempeño de los participantes fuera valorado a partir de un paradigma normativo formal y ofrecieron una reinterpretación de los resultados distinguiendo dos tipos de racionalidad: una racionalidad *instrumental* o *heurística* (r_1) y otra *lógica* o *abstracta* (r_2). La primera –r_1– es una noción heurística de racionalidad ligada a un éxito general, que es generalmente confiable para obtener nuestras metas. La segunda –r_2– emplea estándares que derivan de la lógica, la teoría de la probabilidad y la teoría de la decisión.[12] Evans y Over remiten el origen de esos procesos a dos sistemas cognitivos distintos, S_1 y S_2. Asimismo, consideran que la mayoría del razonamiento que se lleva a cabo mediante los procesos que guían a r_1 es tácito e inaccesible al sujeto. Por tal, desarrollaron una propuesta que dice que la mayor parte del razonamiento que realiza S_1 está guiado por procesos que son implícitos, como lo son aquellos procesos que determinan la relevancia y la atención selectiva y que normalmente nos permiten conseguir nuestras metas. En estos procesos la selección de datos, la relevancia que toman y la influencia de nuestras creencias o conocimientos previos son parte intrínseca de la manera en que esos procesos operan. Evans y Over consideraron legítima la intervención de las creencias en el caso de la racionalidad instrumental, "sugiriendo que es adaptativo que

[12] En escritos más recientes, Evans prefiere utilizar el término *analítico*, ya que ha dejado de considerar correcto sólo apelar a las reglas de la lógica, el cálculo de probabilidades y la teoría de la decisión.

nuestro razonamiento se contextualice automáticamente con el conocimiento previo" (Frankish y Evans 2009, p. 16). Ahora bien, dado que S_2 sólo puede aplicarse a las representaciones que S_1 ha seleccionado, el sesgo se observa cuando i) se excluye información lógica relevante o ii) el procesamiento heurístico incluye información lógicamente irrelevante (Evans 2008, p. 263).

Así, Evans y Over concluyen que lejos de que los resultados reflejen errores y sesgos cognitivos que llevan a los psicólogos cognitivos a la hipótesis de que los sujetos son irracionales, los resultados se pueden tomar como evidencia de que en el sujeto operan diferentes procesos. Los procesos lógicos parecen competir y entrar en conflicto con las respuestas en las que intervienen factores no lógicos y, cuando esto sucede, el razonamiento puede desviarse de los principios normativos deductivos.

Con esta solución de Evans y Over, la respuesta de un sujeto a un cierto problema cognitivo podría ser evaluada como irracional desde la visión de r_2, pero podría ser racional juzgada desde r_1. La idea es que alguien puede ser racional (r_1) en términos de alcanzar metas personales (racionalidad *instrumental*) o ser racional (r_2) en el sentido de conformarse a un sistema normativo como el de la lógica (racionalidad *normativa*). Evans y Over (1996) sostuvieron que la racionalidad instrumental (r_1) no requiere involucrar la racionalidad normativa (r_2), en el sentido de que se sigan *explícitamente* las normas prescritas por un sistema normativo como el de la lógica o la teoría de la probabilidad. En un contexto cotidiano es *instrumentalmente* racional –r_1– que en nuestras tareas de razonamiento influyan nuestras creencias previas, pero desde el punto de vista de r_2 sería irracional (en el sentido normativo de la lógica). Del mismo modo, el buen procesamiento del sistema explícito dará resultados que sean racionales según r_2, pero no necesariamente racionales para r_1. Esto muestra que no existe una relación completamente paralela entre r_1 y el tipo de proceso tácito, y r_2 y el tipo de proceso explícito (*ibid.,* p. 147). Por otra parte, el razonamiento explícito es muchas veces usado para *racionalizar* el comportamiento inconscientemente controlado, pero otras veces es la expresión de un razonamiento que es controlado por el sistema S_2, y el problema es que no hay una manera sistemática de distinguir de dónde procede la respuesta.

En las teorías de *proceso dual* se sostiene que cada proceso exhibe regularmente un diferente grupo de propiedades, de tal manera que las propiedades que corresponden a cada uno de los sistemas forman un haz o conjunto de propiedades co-variantes. Este rasgo de co-variación de propiedades es central para la tesis de que hay dos procesos cognitivos:

los que satisfacen al conjunto de propiedades de S_1 o al conjunto de propiedades de S_2. Dado que ninguna de las propiedades del conjunto depende lógicamente de las otras propiedades, se puede explicar por qué constituyen una colección co-variante de propiedades postulando para cada sistema un conjunto de mecanismos subyacentes que dan sustento a los respectivos procesos (Samuels, 2009, pp. 130-131). Evans ofrece el siguiente listado de rasgos, tomado de diferentes autores, que suelen atribuirse a los sistemas duales cognitivos:

S_1	**S_2**
Automático	Controlado
Rápido	Lento
Capacidad elevada	Baja capacidad
Evolutivamente antiguo	Evolutivamente más reciente
Compartido con otros animales	Distintivamente humano
Inconsciente, preconsciente	Consciente
Paralelo	Secuencial
Conocimiento implícito	Conocimiento explícito
Intuitivo	Reflexivo
Asociativo	Basado en reglas
Contextualizado, pragmático	Abstracto, lógico
Independiente del lenguaje	Dependiente del lenguaje
Independiente de una inteligencia general	Ligado a una inteligencia general

Asimismo Evans sugiere que la definición mínima de los dos sistemas se haga en términos de los tres primeros rasgos y no emplear de entrada el rasgo consciente/inconsciente o preconsciente, con el fin de evitar viciar la discusión. Si bien para algunos existen dos *sistemas* cognitivos o de razonamiento (r_1 y r_2) que pueden operar, uno de manera inconsciente o tácita (r_1) y el otro de manera consciente (r_2), Evans considera que incluso en el proceso de razonamiento r_2 pueden operar procesos tácitos, inconscientes.

Otros agregan, como rasgos distintivos de cada sistema, que el procesamiento se da en paralelo o de forma secuencia, que la capacidad es elevada o baja, o la admisión o no de diferencias individuales. Y quienes ligan la diferencia de rasgos con la cuestión de la arquitectura de la mente añaden, además del par inconsciente o consciente, los rasgos de la evolución antigua o reciente y también que otros animales comparten con los humanos el sistema S_1 pero no S_2 (Evans y Frankish 2009, pp. 33-34). E incluso hay quienes conectan la cuestión evolutiva con diferentes

localizaciones neurológicas, según el tipo de actividad que se esté desempeñando.

En sus escritos más recientes, Evans incluye en el sistema S_1 los procesos en los que intervienen elementos emocionales, aunque no elabora el modo en que lo hacen y tampoco dice si su presencia genera creencias irracionales. Mi sugerencia es que el autoengaño cabe dentro de estas formas y podría incorporarse dentro de las estrategias heurísticas de las que puede valerse el sujeto. Si estoy en lo correcto, queda abierta la posibilidad de que no sean irracionales –valoradas desde S_1–, si se preserva el valor consecuencialista de que por ese medio se consiguió el fin buscado, aunque no de manera consciente.

En los procesos en los que interviene el sistema S_2 también pueden intervenir elementos emocionales, según Evans, pero serían más complejos que los que intervienen en S_1. Pero, de manera similar a lo expresado respecto de los procesos de S_1, cabe preguntarse si, por el solo hecho de que intervinieran en S_2, serían irracionales.

En la actualidad, las diferencias entre las distintas teorías duales se han multiplicado y la dirección que toman muchos autores es la de anclar la distinción en algún rasgo distinto al de los procesos, ya que ahora consideran que en ambos sistemas opera una variedad de procesos. Evans, por ejemplo, incluso postula la existencia de un tercer tipo de procesos, involucrados en disparar la actividad de S_2, y que media entre los dos sistemas, por lo que propone distinguir los sistemas con base en que sólo el sistema S_2 emplea la memoria operativa (*working memory*).[13] El rasgo de dependencia o independencia de una inteligencia general o capacidad cognitiva o memoria operativa se conecta con la idea de que existan o no diferencias individuales. La memoria operativa, según Evans, requiere que haya contenidos y que estos sean proporcionados por diversos sistemas cognitivos *implícitos*, como por ejemplo las representaciones visuales y perceptivas del mundo, los significados extraídos del discurso lingüístico, las memorias episódicas y las creencias recuperadas o rescatadas que son pertinentes en el contexto actual (2009, p. 37).

En escritos más recientes, Evans ha abandonado también la tesis de que sólo en S_1 intervienen las creencias y sostiene que en los dos sistemas

[13] Evans cree que es más factible considerar que S_2 constituye un solo sistema, ya que requiere el uso de un único sistema de memoria operativa, pero no recomienda identificar S_2 con la memoria operativa porque implicaría que un único sistema se sobrecargaría, ya que tendría que ser capaz de llevar a cabo las múltiples tareas asociadas con la memoria operativa (tareas como leer, razonar, planificar, llevar a cabo un aprendizaje explícito, entre otras).

participan las creencias o el conocimiento previo. Tal modificación –con creencias procedentes de dos tipos de memoria, la de largo plazo y la operativa– abre la posibilidad de que se pueda hablar de diferentes formas de autoengaño y que puedan valorarse en cuanto a su irracionalidad con parámetros diferentes. Veamos entonces, cuáles serían éstas.

IV

En esta sección abordo la manera en que Johnston –un destacado defensor de la visión no intencionalista– entiende los procesos mentales, cómo introduce la cuestión de la racionalidad y las consecuencias que deriva para el caso del autoengaño. Posteriormente comparo su propuesta con la que aquí se propone, tomando como plataforma la teoría dual.

Considero que el análisis de Johnston tiene ciertas virtudes y ventajas en comparación con las opciones intencionalistas, al introducir su noción de tropismos mentales. Sin embargo, en el marco de la teoría dual, esta idea puede aprovecharse para que los casos en los que los procesos mentales tácitos que se llevan a cabo en S_1 puedan considerarse racionales, aunque no sean guiados o controlados por el sujeto. Concluyo que son estrategias o estratagemas útiles que el sistema S_1 emplea para obtener ciertos fines. Cuando esto sucede, las transiciones pueden reflejar un patrón de razonamiento heurístico que tenga entre sus premisas deseos del sujeto y cuya conclusión sea una creencia, aunque no sea verdadera, pero que se legitima por los beneficios que aporta. De este mismo modo, en S_1 podría legitimarse también el proceso de autoengaño. Por otra parte, al admitirse en el modelo dual la división de sistemas, se utiliza un modelo de fragmentación como el que suscriben los intencionalistas: cuando con base en la creencia, que en S_2 se asume conscientemente, el sujeto puede racionalizar el proceso que lo llevó a la creencia sin corresponder al proceso que realmente lo llevó a asumirla que, al igual que cuando la creencia es producto de autoengaño, se considerarían irracionales, al igual que lo harían los intencionalistas.

Para entender el funcionamiento de la mente, Johnston sugiere dar cuenta de los procesos mentales, en términos generales, como conexiones entre tipos de estados mentales, mediante la introducción de lo que denomina *tropismos mentales.* Los tropismos mentales son regularidades mentales no accidentales que sirven para un propósito, pero que no son intencionales. El autoengaño pertenece al tipo de casos en los que los deseos y otros intereses del sujeto intervienen en la formación de una creencia pero, debido a la presencia de esos elementos motivacionales, se produce una desviación sub-intencional en la formación de la creencia del

sujeto. Considera que el autoengaño es una instancia de un proceso que tiene un propósito, que responde a ciertos intereses del sujeto, pero es sub-intencional porque el autoengañado no inicia –o guía– el proceso por esos intereses ni por ninguna otra razón. (*op. cit.,* p. 65) Para Johnston, proceso mental sub-intencional involucrado en el autoengaño –y en el pensamiento desiderativo– es aquel que se *explica* por el deseo del sujeto de que se dé *p,* acompañado de la ansiedad de que su deseo de *p* no se satisfaga, ésta se reduce adquiriendo la creencia de que *p.* En el proceso –continúa la explicación–, el sujeto deja de reconocer o soslaya, por cualesquiera medios –como represión, negación o subestimación de datos–, la evidencia que apoya a no-*p.* La sub-intencionalidad del mecanismo de represión reside en que tiene el propósito de reducir la ansiedad del sujeto y que va de acuerdo con el interés del sujeto de lo que desea creer (p. 76).

Johnston compara la no intencionalidad del autoengaño con el movimiento de leer dirigiendo los ojos hacia la parte superior de las letras, lo cual se explica por la función que éstos tienen, para lo que sirven. Escribe: "Desde luego, la explicación de por qué muchos de nosotros usamos de manera no propositiva (*unwittigly*) este método de lectura tiene que ver con el hecho de que nos permite leer más rápido. El método, una vez que damos con él, persiste porque sirve para un propósito; no se emplea intencionalmente para servir a ese propósito" (*ibid.,* p. 86). En el caso de la lectura se trata de un mecanismo disposicional, a pesar de que esta forma de leer haya sido adquirida.

Los tropismos son para Johnston mecanismos causales que constituyen las bases para las conexiones entre diferentes estados mentales, sean tales conexiones racionales o irracionales. Así, en los procesos racionales normales –continúa Johnston– también operan esos mecanismos causales –los tropismos– en los que un estado mental, digamos una creencia, lleva a otro estado mental, otra creencia, y que podemos calificar como un proceso racional si es el caso que la primera creencia constituye una razón para la segunda creencia. Cuando evaluamos que ese proceso es racional –agrega Johnston– no se requiere invocar ningún elemento independiente adicional; el proceso causal es un proceso natural, *ciego,* y es racional si de hecho *se conforma a* un patrón racional. Johnston alude a las *Investigaciones filosóficas* de Wittgenstein, a la sección 201 referente a la muy debatida distinción entre "conformarse a una regla" y "seguir una regla" (véase su nota 31). Según su lectura de esta distinción, Wittgenstein sostiene que en nuestro entendimiento de una regla mostramos que la captamos sin necesidad de que al seguir la regla medie una interpretación. De manera similar, cuando nos proponemos guiarnos por la razón –cuando actuamos intencionalmente–

se da el mismo mecanismo causal de manera ciega. Gracias a mi buena formación instructiva, mi razonamiento se conforma a la forma lógica de *modus ponens* y podría proponerme explícitamente guiar mi pensamiento de acuerdo con él. Esto es, se da en mí un proceso causal que *va* de mi deseo de guiar mi pensamiento de esa manera –mi creencia de que *p* y mi creencia de que si *p* entonces *q* y mi creencia de que el *modus ponens* prescribe que crea *q*–, *a* mi creencia en *q*. Para Johnston, en este razonamiento entre estados mentales el proceso causal es acorde con las transiciones del *modus ponens* y lo único que interviene, en el caso racional e intencional de ajustarse a ese patrón, son la disposición innata, el entrenamiento lógico y el empleo de la capacidad para inhibir operaciones que compiten con ella y que serían irracionales (p. 87).[14]

Johnston hace la conexión entre racionalidad y creencia motivada construyendo un razonamiento práctico en el que el sujeto cree que creyendo *p* hace más factible que *p* se dé, esto es, que tiene el efecto de reducir su ansiedad de que ¬ *p* resulte. Es razonable estar menos ansioso de que ¬ *p* si uno se convence de que *p*. Para Johnston esta explicación la hace el teórico para hacer inteligible la conducta del sujeto; sin embargo, piensa que en el sujeto esta conexión racional no interviene en la generación de la creencia (*ibid*., p. 74). Al no tomarse en cuenta la evidencia en el proceso mental del autoengaño, da pie para que Johnston lo ubique al margen de la racionalidad y del razonamiento. Para Johnston –al igual que para el intencionalista– racionalidad e intencionalidad vienen juntas. Hasta aquí la propuesta de Johnston.

En suma, considero que las virtudes de la propuesta de Johnston radican en que los procesos mentales en general: a) se sustentan en tropismos o mecanismos causales; b) en la admisión de que hay procesos sub-intencionales, pero que tienen un propósito o función, y c) en que estos mecanismos no son racionales ni irracionales. Sin embargo, estoy en desacuerdo con: a) la conexión que asume entre intencionalidad y racionalidad; b) su rechazo a la mediación de un razonamiento en el caso

[14] Saunders y Over (2009), en tanto defensores de una versión de la teoría de proceso dual, también apelan a esta distinción wittgensteiniana. Para estos autores, en relación con la racionalidad instrumental, se pueden obtener las metas de forma confiable de manera implícita (conformándose a reglas) o de manera explícita (siguiendo reglas). La diferencia entre la interpretación de Johnston y la de estos autores radica en que la distinción para Johnston se relaciona con los tropismos, mientras que para estos autores se relaciona con las reglas. Agregan que en el primer caso se requieren habilidades que se encuentran asociadas a módulos más específicos, mientras que en el segundo se requiere el entendimiento de condicionales explícitos y, para poder hacer inferencias explícitas se requieren habilidades más generales (p. 138).

del autoengaño y, por tanto c) identificarlo con el mecanismo causal subyacente. El modelo dual me sirve para sustentar estos desacuerdos.

Con la tesis de la racionalidad dual se rompe la articulación entre la intencionalidad y la racionalidad, que me abre la posibilidad de considerar que el autoengaño puede involucrar un razonamiento heurístico implícito o tácito. Veamos cómo.

Si existen dos tipos de racionalidad, una instrumental y otra normativa, la primera se evalúa por sus consecuencias y si funciona la estrategia como medio para obtener un fin, entonces es racional. En este sentido es razonable que el estar menos ansioso con respecto a que se dé lo que temo, lleve a la adopción de la creencia contraria. Se trata del mismo razonamiento práctico que, según Johnston, emplea el teórico para hacer inteligible la conducta del sujeto; pero en el sistema S_1 este razonamiento podría integrarse en la forma en la que opera y procesa información. En S_1 se lleva a cabo un razonamiento tácito, que puede ser racional si la estrategia adoptada fue eficaz y lo mismo podríamos decir del proceso del autoengaño. Uno de los beneficios que se obtienen en la forma de procesamiento de S_1 es que requiere poco esfuerzo cognitivo, lo cual permite que se optimicen nuestros recursos. La racionalidad que se asocia a S_1 y sus rasgos, de algún modo se reflejan en el razonamiento con esas mismas características.

En S_2 se procede con base en el tipo de racionalidad $-r_2-$, que no requiere evaluarse por sus consecuencias, excepto si se adopta conscientemente la misma estrategia heurística en la que se basa S_1. El razonamiento heurístico que utiliza el sistema S_2 a) puede guiar al sujeto para obtener cierto resultado o b) puede utilizarlo para corregir o inhibir el resultado que recibe de S_1. Pero si no se interfiere o inhibe el resultado que arroja S_1, la falta sería de S_2, no de S_1, ya que estaría fallando en una de sus funciones. Si se diera un proceso de autoengaño en S_2, en a) es difícil ver cómo podrían evitarse el tipo de paradojas en las que el sujeto intencionalmente asume una creencia que sabe o cree que es falsa. El proceso sería irracional, ya que el autoengaño llevaría a que la intención del sujeto de guiarse por ese proceso se abortaría. En b), no inhibir el resultado que emite S_1, se mostraría un tipo de falla del sujeto para modificar ese resultado y sería irracional si va en contra del razonamiento explícito que racionalmente lo lleva a la respuesta contraria.

El ejemplo que pone Johnston en relación con el *modus ponens*, en el contexto de la teoría dual de razonamiento, podría ubicarse dentro de los procesos racionales de S_1 o de S_2, según se emplee implícita o explícitamente. Sólo en el último caso las premisas aparecen explícitamente en el razonamiento. En el caso de S_1 están implícitamente, en tanto que

el patrón causal concuerda con el patrón de *modus ponens*, y en ese sentido también estaría implícitamente el razonamiento. Me parece que el proceso de autoengaño, cuando en S_1 hay una concordancia implícita con el patrón de *modus ponens*, podría evaluarse como racional pero el proceso de S_2 se evaluaría como irracional. Tendríamos que considerarlo irracional, al igual que si obtuviéramos creencias por hipnosis.

Para el intencionalista, si en el autoengaño intervienen elementos como los deseos, emociones, sentimientos e intereses, se estaría violando otro principio de racionalidad según el cual las creencias no se pueden manipular a voluntad sin producir creencias incoherentes o irracionales. Sin embargo, en una teoría dual como la que aquí se suscribe, puede no ser irracional, pues también posee características que nos permiten conceder que el autoengaño, puede permitir que intervengan legítimamente, además de las creencias, los elementos motivacionales y mecanismos psicológicos, esto es, sin convertirlo en un proceso irracional. Para ser racional bastaría con que con la activación del proceso alcanzara su propósito.

Por otra parte, si bien tanto para Johnston como para los defensores de la teoría dual los procesos cognitivos se basan o apoyan en mecanismos causales, la diferencia radica en que para él la plataforma general para discutir los procesos mentales cognitivos se funda en los mismos mecanismos: los tropismos mentales racionales constituyen "sólo una de las formas adaptativas que los procesos mentales pueden tomar" (p. 89). Para los dualistas, en contraste, a cada sistema le corresponden sus propios mecanismos. En la sección anterior mencioné cómo Samuels reconstruye el argumento para explicar que algunas de las propiedades que pertenecen a cada uno de los sistemas co-varían, postulando diferentes mecanismos.

Asimismo, Johnston no acepta en su propuesta que haya algún tipo de división o de fragmentación del sujeto o de los sistemas, y en eso la teoría dual comparte ciertos elementos con la postura intencionalista y la manera en que en ellos un subsistema se caracteriza por llevarse a cabo de manera inconsciente y el otro de manera consciente. Al no aceptar la división y admitir que hay funciones sub-intencionales, a Johnston no le queda más que identificarlas con los tropismos o mecanismos causales. En una teoría dual cabe la posibilidad de asociar los procesos sub-intencionales a las formas particulares en que los procesos se llevan a cabo en el sistema S_1. El mecanismo –de orden computacional– que subyace a este sistema, aunque limitado en recursos, le da al mismo un gran poder o capacidad cognitiva, consiguiendo que procesos de información muy complejos se realicen de una manera rápida, y donde focalizarse en o seleccionar cierta información juega un papel clave. Pero esta selección forma parte de la operación misma del sistema S_1.

En conclusión, no comparto con Johnston o Lazar (1999) la idea de que deseos o emociones causen parcialmente la creencia producto del autoengaño, que este deba entenderse como un vínculo *causal* y por tanto fuera de cualquier evaluación en la dimensión del par racional/irracional y fuera de cualquier tipo de razonamiento. En cambio, en la línea que sugieren los teóricos duales, propongo que cuando se lleva a cabo un proceso de autoengaño, en S_1 este puede estar mediado por un razonamiento práctico inconsciente. Así, el patrón del autoengaño puede, en ciertas circunstancias, involucrar un tipo de razonamiento práctico y, como tal, ser evaluado como racional o irracional, según se logre o no su propósito. Me percato que esto tiene un costo: sacrificar la exactitud, la exhaustividad en la consideración de posibilidades o alternativas e incluso sacrificar la verdad.

Si admitimos la existencia de dos procesos de razonamiento, se podría defender que ambas posiciones –la intencionalista y la no intencionalista– son parcialmente correctas. Por una parte, en la idea misma de un modelo dual se está suscribiendo algún tipo de fragmentación o división, que al acompañarlo de dos formas de razonamiento, permite que cambie la valoración del autoengaño en cuanto a su racionalidad según se explique bajo una u otra de esas formas de razonamiento. No asume, como lo hace el intencionalista, que la valoración del autoengaño es siempre irracional, ya que no se compromete con la tesis de que el sujeto de alguna manera *guía* o *controla* el proceso. De acuerdo con lo que sostiene el no intencionalista, la capacidad de autoengaño puede formar parte de la batería de estrategias heurísticas que corresponden a S_1 y que hemos desarrollado para lidiar con situaciones que se dan en relación con otras personas o con nuestro medio ambiente. Según cuál respuesta logre imponerse y de qué sistema proceda, o si se logra que los sistemas interactúen coordinadamente, dependerá la legitimidad del recurso del autoengaño y su valoración en cuanto a la racionalidad. Tendremos que decidirlo caso por caso.

En el caso del autoengaño en el nivel consciente, la forma de operar de S_2 supone una interacción entre los dos niveles: la persona tendría una creencia de segundo orden (consciente) que no corresponde a una creencia de primer orden (inconsciente); o a la inversa, tendría una creencia de primer orden, pero no asiente a ella (no se forma una creencia de segundo orden). Si no hay concordancia en las creencias de los dos sistemas, entonces la creencia que asume conscientemente el sujeto sería producto de un autoengaño. Para evitar que este desfase entre la creencia de primer orden y la de segundo sea un simple caso de una creencia falsa y tenga la característica específica de un caso de autoengaño, podría

atribuírsele alguna de las siguientes situaciones: a) que la activación de alguno de los mecanismos de S_1 produce una respuesta que no es asumida, o corregida, por S_2, y b) que la creencia que surge de S_2 (la creencia a la que la persona asiente) es racionalizada por el sujeto. Empero, en ninguno de los dos casos el sujeto se percata de que existe un conflicto de creencias. En el caso normal se da la misma respuesta en los dos sistemas, no hay conflicto pero se llega a esta por caminos diferentes y con distintos recursos, aunque por el resultado no podemos determinar de qué sistema procedió ya que el sistema S_2 no sólo utiliza procesos de manera consciente. Sin embargo, si se acepta que en el autoengaño opera una forma de sub-intencionalidad que no involucra un ejercicio autorreflexivo, se perderá irremisiblemente parte de su misterio y de su carácter paradójico y de este modo también abre la posibilidad de recuperar su función natural en nuestra vida psicológica y social.

Bibliografía

Barnes, A. (1997), *Seeing through Self-Deception*, Nueva York, Cambridge, Cambridge University Press.

Botterill, G. y P. Carruthers (1999), *The Philosophy of Psychology*, Cambridge, Cambridge University Press.

Carruthers, P. (2009), "How do We Know our Own Minds: The Relationship Between Mindreading and Metacognition", *Behavioral and Brain Sciences*, 32, pp. 121-138.

Cherniak, C. (1986), *Minimal Rationality*, Cambridge, Mass., The MIT Press.

Davidson, D. (1974), "Psychology as Philosophy", en Davidson, D. (1980), *Essays on Actions and Events*, Oxford, Clarendon Press. Versión en español: Hansberg, O., Robles, J. A. y Valdés, M. (trads.) (1995), *Ensayos sobre acciones y sucesos*, México-Barcelona, Instituto de Investigaciones Filosóficas-UNAM/Crítica.

——— (1986), "Deception and Division", en Elster, J. (comp.), *The Multiple Self*, Nueva York, Cambridge University Press, pp. 79-92. Versión en español: Hansberg, O. (trad.) (1992), "Engaño y división", en *Quinto Simposio Internacional de Filosofía*, vol. 2, México, Universidad Nacional Autónoma de México, pp. 85-101.

Elster, J. (1999), *Alchemies of the Mind: Rationality and the emotions*, Cambridge, Cambridge University Press.

Epstein, S. (1994), "Integration of the Cognitive and Psychodynamic Unconscious", *American Psychology,* 49, pp. 709-742.

Evans, J. (2008), "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition", *Annual Review of Psychology*, 59, pp. 255–278.

Evans, J. (2009), "How Many Dual-Process Theories do We Need? One, Two or Many?", en Evans, J. y Keith, F. (eds.) (2009), *In Two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press.

——— (2010), *Thinking Twice: Two Minds in One Brain*, Oxford, Oxford University Press.

Evans, J. y Keith, F. (eds.) (2009), *In Two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press.

Evans, J. y Over, D. E. (1996), *Rationality and Reasoning*, Hove, U. K., Psychology Press.

Gilovich, T., Griffin, D. y Kahneman, D. (eds.) (2002), *Heuristics and Biases: The Psychology of Intuitive Judgement*, Nueva York, Cambridge, University Press.

Goldman, A. (1986), *Epistemology and Cognition*, Cambridge, MA, Harvard University Press.

Hassin, R. R., Uleman, J. S. y Bargh, J. A. (eds.) (2005), *The New Unconscious*, Oxford, Oxford University Press.

Hauser, M. (1996), *The Evolution of Communication*, Cambridge, MA, The MIT Press.

Johnston, M. (1988), "Self-Deception and the Nature of Mind", en McLaughlin, B. y Rorty, A. O. (eds.), *Perspectives on Self-Deception*, Berkeley, University of California Press.

Kahneman, D., y Tversky, A. (1982), "On the Psychology of Prediction", en Kahneman, D., Slovic, P. y Tversky, A. (eds.), *Judgement Under Uncertainty: Heuristics and Biases,* Nueva York, Cambridge University Press.

Krebs, J., y Dawkins, R. (1984), "Animal Signals: Mind-Reading and Manipulation", en Krebs, J. y Davies, N. (eds.), *Behavioural Ecology*, Sunderland, MA, Sinauer Associates, pp. 380-402.

Krebs D. L., y Denton, K. (1997), "Social Illusions and Self-Deception: The Evolution of Biases in Person Perception", en Simpson, J. A. y Hendricks, D. T. (eds.), *Evolutionary social psychology*, Mahwah, NJ, Lawrence Erlbaum Associates.

Krebs, D. L., Denton, K. y Higgins, N. (1988), "On the Evolution of Self-Knowledge and Self-Deception", en McDonald, K. (ed.), *Sociobiological Perspectives on Human Behavior*, Nueva York, Cambridge University Press, pp. 142-179.

Lazar, A. (1999), "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence'", *Mind*, 108 (430), pp. 265-290.

Lieberman, M. D. (2003), "Reflective and Reflexive Judgment Processes: A Social Cognitive Neuroscience Approach", en Forgas, J. P.,

Williams, K. R. y von Piel, W. (eds.), *Social Judgments: Implicit and Explicit Processes*, Nueva York, Cambridge University Press.

Lieberman, M. D. (2009), "What Zombies Can't Do: A Social Cognitive Neuroscience Approach to the Irreducibility of Reflective Consciousness", en Evans, J. y Keith, F. (eds.) (2009), *In Two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press.

Lockard, J. S., y Paulhus, D. I. (eds.) (1988), *Self-Deception: An Adaptive Mechanism?*, Englewood Cliffs, NJ, Prentice Hall.

Mele, A. (1987), *Irrationality: An Essay on Akrasia, Self-Deception, Self-Control,* Oxford, Oxford University Press.

——— (1999), "Twisted Self-Deception", *Philosophical Psychology,* 12, pp. 117-137.

——— (2000), "Self-Deception and Emotion", *Consciousness and Emotion,* 1, pp. 115-139.

——— (2001), *Self-Deception Unmasked*, Princeton, Princeton University Press.

Piatelli-Palmarini, M. (2005), *Los túneles de la mente. ¿Qué se esconde tras nuestros errores?,* Pons, M. (trad.), Barcelona, Crítica.

Sahdra, B., y Thagard, P. (2003), "Self-Deception and Emotional Coherence", en Thagard, P. (2006), *Hot Thought: Mechanisms and Applications of Emocional Cognition*, Cambridge, Mass., The MIT Press, cap. 13.

Samuels, R. (2009), "The magical number two, plus or minus: Dual-process theory as a theory of cognitive kinds", en Evans, J. y Keith, F. (eds.) (2009), *In Two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press, pp. 129-146.

Saunders, C. y Over, D. E. (2009), "In Two Minds About Rationality?", Evans, J. y Keith, F. (eds.) (2009), *In Two Minds: Dual Processes and Beyond*, Oxford, Oxford University Press.

Sturm, T. (2007), "Self-Deception, Rationality and the Self", *Teorema*, XXVI (3), Ediciones KRK, Oviedo, España, pp.73-91.

Trivers, R. (1971), "The Evolution of Reciprocal Altruism", *Quarterly Review of Biology*, 46, pp. 81-91.

Williams, B. (1973), "Deciding to Believe", en *Problems of the Self*, Cambridge, Cambridge University Press, pp. 136-151. 1a. ed. de Kiefer, H. E. y. Munitz, M. K (comps.) (1970), *Language, Belief and Metaphysics*, Albany, State University of New York Press, pp. 95-111. Versión en español: "Decidirse a creer", en Helguera, J. M. G. (trad.) (1986), *Problemas del yo*, México, Instituto de Investigaciones Filosóficas-UNAM, pp. 181-200.