

## Autoengaño, sistemas de creencias y errores en el autoconocimiento

*Gustavo Fernández Acevedo*  
*Universidad Nacional de Mar del Plata*

**Resumen:** Richard Holton ha cuestionado la idea usual de que el autoengaño consiste en un engaño *por* el sí mismo, y ha propuesto en su lugar que la caracterización de este fenómeno debe incluir, como condición necesaria, la tesis de que el autoengaño es un engaño *acerca* del sí mismo. Se defiende aquí la afirmación de que tal requisito no es necesario, sobre la base de dos características centrales del autoengaño: su carácter temporal y su tendencia a la multiplicación. Asimismo, se esboza una condición alternativa respecto del conjunto de creencias involucrado en el autoengaño.

**Palabras clave:** autoengaño; sí mismo; autoconocimiento; sistemas de creencias; temporalidad

**Abstract:** “Self-Deception, Beliefs Systems and Self-knowledge’s Errors”. Richard Holton has criticized the common idea that self-deception is deception *by* the self, and suggested it is rather deception *about* the self; self-deception must include necessarily erroneous beliefs about the self. In this article I claim that this condition is not necessary, based on two central traits of self-deception: its temporal character and its bound to multiplication. In addition, I suggest an alternative condition in relation to the beliefs system implied in self-deception.

**Key words:** self-deception; self; self-knowledge; belief systems; temporality

I

En las últimas cinco décadas el fenómeno del autoengaño ha sido objeto de una creciente atención no solo por parte de la disciplina en la cual surgió como problema teórico, la filosofía, sino también por parte de un espectro de disciplinas científicas que incluye la biología evolucionista, la psicología, las neurociencias y las ciencias sociales.

Esta ampliación del campo disciplinar trajo aparejada, a su vez, una proliferación de interrogantes teóricas que se adicionaron a los clásicos planteos filosóficos sobre el tema (entre ellas, por ejemplo, cómo es posible que alguien se engañe intencionalmente a sí mismo, cómo puede alguien sostener creencias contradictorias y cuál es la responsabilidad moral concerniente al autoengaño). Esos planteos recientes incluyen, entre otros, preguntas sobre las posibles ventajas adaptativas<sup>1</sup> del autoengaño<sup>2</sup>, sobre su potencial contribución con el bienestar y la felicidad<sup>3</sup>, sobre el producto del autoengaño, esto es, si el resultado del proceso que conduce a este estado es una creencia u otra

---

<sup>1</sup> Con la expresión “ventajas adaptativas” estamos haciendo referencia a los posibles beneficios que, en un sentido evolucionista, podría haber tenido el autoengaño para la supervivencia y la aptitud de aquellos organismos que poseyeran la capacidad para la producción de este fenómeno.

<sup>2</sup> Al respecto, cf. Trivers, R., “The Elements of a Scientific Theory of Self-Deception”, en: *Annals of the New York Academy of Sciences*, 907 (2000), pp. 114-131; Trivers, R., *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*, Nueva York: Basic Books, 2011; Van Leeuwen, D. S., “The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon”, en: *Philosophical Psychology*, XX, 3 (2007), pp. 329-348; Van Leeuwen, D. S., “Finite Rational Self-Deceivers”, en: *Philosophical Studies*, 139 (2008), pp. 191-208; von Hippel, W. y R. Trivers, “The Evolution and Psychology of Self-Deception”, en: *Behavioral and Brain Sciences*, 34 (2011), pp. 1-56; Kido Lopez, J. y M. J. Fuxjager, “Self-Deception’s Adaptive Value: Effects of Positive Thinking and the Winner Effect”, en: *Consciousness and Cognition*, 21 (2012), pp. 315-324.

<sup>3</sup> Cf. Taylor, S. y J. Brown, “Illusion and Well-Being: A Social Psychological Perspective on Mental Health”, en: *Psychological Bulletin*, CIII, 2 (1988), pp. 193-210; Taylor, S. y J. Brown, “Positive Illusions and Well-Being Revisited. Separating Fact From Fiction”, en: *Psychological Bulletin*, CXVI, 1 (1994), pp. 21-27; Colvin, R. C. y J. Block, “Do Positive Illusions Foster Mental Health? An Examination of the Taylor and Brown Formulation”, en: *Psychological Bulletin*, CXVI, 1 (1994), pp. 3-20; Van Leeuwen, D. S., “Self-Deception Won’t Make You Happy”, en: *Social Theory and Practice*, XXXV, 1 (2009), pp. 107-132.

clase de estado intencional<sup>4</sup> y sobre la posibilidad de que el autoengaño sea un fenómeno cultural, más que uno intrapsíquico<sup>5</sup>.

En un artículo relativamente reciente, Richard Holton<sup>6</sup> agrega una pregunta menos frecuente que las anteriores, pero que, además de su interés teórico intrínseco, posee implicaciones de gran alcance para algunas teorías explicativas sobre el fenómeno referido. En particular, Kido Lopez y Fuxhager<sup>7</sup> emplean las conclusiones de Holton como parte de un argumento tendiente a demostrar que el autoengaño usualmente conduce a la formación de creencias positivas acerca de uno mismo. La adopción de la tesis de Holton les permite sostener que el autoengaño, independientemente de cuáles hayan sido los mecanismos de la evolución que le dieron origen (esto es, independientemente de si surgió como una adaptación o como un *spandrel* o subproducto no adaptativo), conduce a un beneficio en la aptitud para los organismos que poseen tal capacidad. En caso de ser correcto, en consecuencia, el argumento de Holton constituiría una contribución importante a las polémicas relativas a la explicación evolucionista de la existencia del autoengaño.

Holton observa que las perspectivas ortodoxas acerca del papel del sí mismo [*self*] en el autoengaño afirman que es el sí mismo el que se engaña; esto es, se trata de un engaño *por* el sí mismo. Su interés es defender una idea distinta, en particular, la idea de que el autoengaño implica un engaño *acerca* de sí mismo. Su tesis será, en consecuencia, que la caracterización del autoengaño debe incluir, como condición necesaria, la existencia en el sujeto autoengañado de creencias falsas respecto de sí mismo; esto es, el autoengaño requiere, de manera esencial, una falla en el autoconocimiento.

El planteo de Holton parece indudablemente legítimo al menos en un aspecto. Pese a que las caracterizaciones del autoengaño suelen incluir como condición necesaria la presencia de una creencia falsa en el sistema de creencias

<sup>4</sup> Cf. Funkhouser, E., "Do the Self Deceived Get what they Want?", en: *Pacific Philosophical Quarterly*, 86 (2005), pp. 295-312; Van Leeuwen, D. S., "The Product of Self-Deception", en: *Erkenntnis*, LXVII (2007), pp. 419-437.

<sup>5</sup> Cf. Harré, R., "The Social Context of Self-Deception", en: McLaughlin, B. y A. Oksenberg-Rorty (eds.), *Perspectives on Self-Deception*, Berkeley: University of California Press, 1988, pp. 364-379; Lewis, B., "Self-deception: A Postmodern Reflection", en: *Journal of Theoretical and Philosophical Psychology*, XVI, 1 (1996), pp. 49-66; Clegg, J. W. y L. Moissinac, "A relational theory of self-deception", en: *New Ideas in Psychology*, 23 (2005), pp. 96-110.

<sup>6</sup> Cf. Holton, R., "What is the Role of the Self in Self-Deception?", en: *Proceedings of the Aristotelian Society*, 101 (2000), pp. 53-69.

<sup>7</sup> Cf. Kido Lopez, J. y M. J. Fuxjager, *o.c.*

del individuo autoengañado<sup>8</sup>, resulta muy plausible la suposición de que esta creencia falsa no puede ser única, sino que debe estar acompañada por un conjunto estructurado de creencias falsas coherentes con ella. En este sentido, la suposición de que este sistema de creencias incluye creencias falsas sobre el sí mismo se torna muy plausible. Sin embargo, en este artículo intento mostrar, contra la pretensión de Holton, que el autoengaño no requiere esencialmente de creencias erróneas sobre uno mismo; más aún, argumentaré que es dudoso que requiera de creencias sobre uno mismo en absoluto, excepto en los casos en los que la creencia principal del autoengaño sea una creencia *sobre* uno mismo. No obstante, es posible dar sentido a la idea de un conjunto de creencias falsas como condición necesaria para el autoengaño.

En la siguiente sección, presentaré una reconstrucción del argumento de Holton referido a la falla en el autoconocimiento como requisito necesario para el autoengaño. En la tercera sección, intentaré mostrar el porqué, a mi modo de ver, fracasa en su intento de demostrar que la falla en el autoconocimiento constituye una condición necesaria para la caracterización del autoengaño, y esbozaré una posible condición alternativa que retenga la idea de que el autoengaño implica no simplemente una creencia falsa, sino un sistema de creencias falsas organizado en torno a una creencia central.

## II

Como hemos anticipado, Holton señala que la posición ortodoxa sostiene que la idea común acerca del autoengaño es que se trata de un engaño *por* el sí mismo: el sí mismo actúa para engañarse a sí mismo, es decir, es el autor de su propio engaño. Su objetivo es explorar una idea opuesta: el sí mismo es más bien el *objeto* [*subject matter*] del engaño, esto es, concierne más al engaño del sí mismo *acerca* de sí mismo que al engaño del sí mismo *por* el sí mismo. Bajo este enfoque, lo que es incorrecto cuando estamos autoengañados es que carecemos de autoconocimiento o, más precisamente –dado que podemos carecer de auto-

---

<sup>8</sup> O, como mínimo, una creencia que es incompatible con la evidencia a disposición del agente. Véase al respecto la siguiente caracterización: “un agente está en un estado de autoengaño si y solo si sostiene una creencia que a) es contraria a lo que sus normas epistémicas, en conjunción con la evidencia disponible, usualmente dictarían y b) un deseo de obtener cierto estado de cosas, o de poseer cierta creencia, producen una diferencia causal de modo que la creencia es sustentada en un modo epistémicamente ilegítimo” (Van Leeuwen, D. S., “The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon”, p. 332).

conocimiento sin estar en un error–, que lo que es incorrecto en el autoengaño es que tenemos creencias *falsas* acerca de nosotros mismos. Específicamente, el autoengaño requiere de creencias falsas acerca del tipo de cuestiones que, si uno estuviera en lo correcto, constituirían al autoconocimiento.

Sin embargo, la formulación anterior abre paso a una interrogante que Holton hace explícita: una cosa es decir que los casos de autoengaño típicamente involucran un error acerca del sí mismo, y otra es decir que tal tipo de error es *esencial* para el concepto. Esta segunda afirmación es la que el autor quiere investigar.

¿Qué implicaría exactamente la segunda afirmación? La posición más débil es que el concepto ordinario de autoengaño es ambiguo: puede significar tanto engaño por el sí mismo como acerca del sí mismo. Sin embargo, lo que el autor quiere afirmar es que el error acerca del sí mismo es *necesario* para el autoengaño. Tal afirmación es polémica: el esposo traicionado por su esposa pero confiado en su fidelidad, ¿no está engañado acerca de ella, pero no acerca de sí mismo? Holton sugerirá que aun aquí se requiere un error acerca del sí mismo si estamos ante un caso de autoengaño: la víctima debe estar en un error acerca de cómo justifica sus creencias. En este punto puede surgir la objeción de que el autoengaño va a ser tratado como un error acerca del sí mismo y que el error acerca de la propia esposa apenas contaría como tal, incluso si es apuntalado por errores acerca del sí mismo. Podría parecer, dice el autor, que está ofreciendo ahora una perspectiva disyuntiva mucho menos atractiva: o bien el autoengaño requiere un error acerca de sí mismo, o bien un error acerca de algo más que es apuntalado por un error acerca de sí mismo. Holton sostiene que esta apariencia de heterogeneidad es engañosa. El punto crucial, observa, es que “autoengañado”, como “engañado”, no adopta un complemento oracional. No es posible decir que alguien está autoengañado *de que* p; más bien, cuando uno está autoengañado, está engañado *acerca* de algo. Y lo que reemplaza a “algo” es una frase sustantiva: uno está engañado acerca de, digamos, los propios motivos, o los propios deseos, o las cualidades de nuestros hijos. El autoengaño concierne a un *asunto*, no a una creencia singular.

Lo anterior no implica negar el rol central de la creencia falsa: si una persona está autoengañada acerca de un asunto, entonces algunas de sus creencias acerca del asunto son creencias falsas. La propuesta de Holton es que algunas de esas creencias falsas, particularmente algunas que son lo suficientemente centrales, deben ser creencias acerca del sí mismo. Las creencias del marido engañado acerca del estatus de su propia evaluación de la fidelidad de

su esposa son parte de su particular conjunto de creencias acerca de la fidelidad de su esposa. Es verdad, observa Holton, que una vez que se dice que creencias particulares son el resultado de un autoengaño, parecería que somos forzados a volver a una caracterización disyuntiva: o esa creencia, o una creencia en la cual aquella está basada, son acerca del sí mismo. No obstante, señala Holton, no debe pensarse que tal enfoque es *fundamentalmente* disyuntivo.

Tal concepción formará la base del argumento de Holton de que el error acerca del sí mismo es *necesario* para el autoengaño. Surge aquí la pregunta relativa a si podría argumentarse que el error acerca del sí mismo es necesario y *suficiente* para el autoengaño. Si bien considera que hay razones legítimas que hacen que la tentación de afirmar esto sea fuerte, Holton observa que es necesario resistirse a ella. La razón para esto es que uno puede estar en un error acerca de sí mismo sin estar autoengañado. Apelando a que puede ocurrir que alguien está en un error acerca de sí mismo, pero no por ello es culpable de dicho error, Holton vuelve a la idea de que el autoengaño requiere que el sí mismo sea el engañador; idea llamativa que podría formularse llanamente como: *uno es el que se engaña a sí mismo*. Propone dos respuestas ante esta idea. La primera consiste simplemente en reconsiderar la afirmación acerca de una condición necesaria. En la medida en que haya un modo de dar un sentido coherente a la idea de que uno puede (quizás no intencionalmente) engañarse a uno mismo, considera que es mucho más plausible un enfoque de nuestro concepto común de autoengaño si le adicionamos una condición necesaria que concierna al asunto: debe involucrar engaño *acerca* del sí mismo. La segunda respuesta, señala, es más especulativa. Consiste en que, una vez que la idea de que el autoengaño involucra una carencia de autoconocimiento ha sido reubicada en el centro de la escena, es posible poner la concepción del yo como un engañador genuino nuevamente en su lugar apropiado. Esto es un intento de explicar cómo tal carencia de autoconocimiento es posible, y por qué deberíamos ser considerados responsables por ella.

Holton comienza su defensa de la primera respuesta señalando lo que cree que es incorrecto en el enfoque de Mele acerca del autoengaño. Describe las cuatro condiciones que Mele<sup>9</sup> sugiere como conjuntamente suficientes para adquirir o mantener una creencia de que  $p$ <sup>10</sup>, y argumenta que tales condi-

---

<sup>9</sup> Cf. Mele, A., *Self-deception Unmasked*, Princeton: Princeton University Press, 2001.

<sup>10</sup> “1. La creencia de que  $p$  que adquiere  $S$  es falsa. / 2.  $S$  trata datos relevantes, o al menos aparentemente relevantes, respecto del valor de verdad de  $p$  de un modo motivacionalmente sesgado. / 3. Este tratamiento sesgado es una causa no desviante de que  $S$  adquiera la

ciones no son suficientes. Para ello propone inicialmente el ejemplo de Jean Marie, un racista que está convencido de la superioridad de los blancos sobre los negros y los árabes<sup>11</sup>. Holton considera que Jean Marie, si bien cumple con las cuatro condiciones de Mele, no está obviamente autoengañado. La razón que presenta para afirmar esto es que sus falsas creencias son acerca del tipo erróneo de cosas; no está cometiendo *obviamente* un error acerca de sí mismo. Sin embargo, en un análisis más cuidadoso, parecería que sí lo comete, por ejemplo, al pensar de sí mismo que sopesa la evidencia de modo equilibrado, cuando en realidad no lo hace.

Si bien el caso examinado parecería demostrar que no es posible hallarse en un estado de autoengaño sin a la vez poseer creencias erróneas acerca de uno mismo, Holton presenta otro caso del que afirma que satisface las cuatro condiciones de Mele, pero no constituye autoengaño. Este es el ejemplo de Catherine, quien se ha postulado para varios empleos recientemente, y no ha tenido éxito en ningún caso. Ha quedado terriblemente decepcionada en cada oportunidad. En cada ocasión invirtió una gran cantidad de tiempo pensando en el trabajo; esto tuvo como resultado el imaginar lo que sería obtenerlo, y se sintió devastada al no lograrlo. Ahora ha decidido postularse para otro empleo. Piensa que es claramente mejor que su empleo actual, de otro modo

---

creencia de que  $p$ . / 4. El cuerpo de datos poseidos por  $S$  en ese momento provee mayor justificación para  $\sim p$  que para  $p$ " (Mele, A., *Self-Deception Unmasked*, pp. 50-51).

<sup>11</sup> Si bien es una observación tangencial frente los fines de este artículo, me parece conveniente señalar que el ejemplo del racista, así como otros ejemplos análogos (el del fanático religioso o del adherente entusiasta y acritico a ideas o regímenes políticos), no es, a mi modo de ver, el mejor caso que puede examinarse para determinar qué condiciones debe cumplir el autoengaño. En efecto, el caso del racista incluye algunos rasgos que complican en buena medida el análisis, e incluso la propia atribución de autoengaño. En primer lugar, no se trata de un conocimiento sobre estados o sucesos singulares (como el propio estado de salud). Se trata, más bien, de afirmaciones sobre regularidades que exceden ampliamente el dominio de los hechos que nos incumben directamente (si bien esta es una categoría vaga, sin duda tiene casos claros de aplicación, como los referidos a nuestro estado de salud o a la fidelidad de nuestra pareja). En segundo lugar, no se trata de creencias sobre las cuales se pueda tener evidencia directa (como sí se puede, por ejemplo, respecto a la creencia sobre la infidelidad de la pareja). La mayor parte de la evidencia pertinente para la formación de creencias que se refieren a colectivos (como los grupos étnicos) será una evidencia mediatizada (por ejemplo, por terceros y/o por medios de comunicación). Esto hace mucho más difícil establecer, en un sentido normativo, cuál es la evidencia pertinente a disposición del agente a partir de la cual puede sustentarse la creencia rechazada por el agente. Tercero, tales ejemplos comparten el atributo de que muy a menudo son compartidos por colectivos más o menos extensos; no se trata de meros procesos de autoengaño individual. Estas tres características, no compartidas por muchos casos aparentemente típicos de autoengaño, hacen que tales ejemplos, a nuestro modo de ver, dificulten notoriamente el análisis y agreguen complicaciones adicionales innecesarias.

no postularía. Pero, esta vez ha resuelto no pensar muy profundamente acerca de cómo es el trabajo, o examinar la evidencia que posee, al menos no hasta que le hagan una oferta. Sabe que, en la medida en que piense más acerca del trabajo, hay alguna posibilidad de que cambie su opinión acerca de los méritos de este. No obstante, Catherine considera que la posibilidad de error es un riesgo aceptable para mantener su ecuanimidad. Resulta luego que ella estaba radicalmente equivocada acerca del trabajo. Es un trabajo terrible, mucho peor que su trabajo actual, como una escasa reflexión adicional le habría mostrado.

Holton considera que este ejemplo no constituye autoengaño, sino un caso de “ceguera deliberada” [*willfully blind*]. Catherine no está autoengañada; simplemente, y bastante racionalmente, rehúsa echar una mirada sobre cierta evidencia que posee. Sin embargo, argumenta, el ejemplo de Catherine podría ser convertido en un caso de autoengaño si ella fuera menos conciente de su propia estrategia, si de hecho evitara encontrarse con la verdad acerca del trabajo, pero lo negara vehemente y sinceramente frente a alguien que sugiriera que esa es la realidad. En este caso (y en una variante alternativa que presenta), Holton sostiene que estaríamos en presencia de autoengaño. Lo que hace la diferencia entre ceguera deliberada y autoengaño es la ausencia o presencia en Catherine de creencias falsas acerca de sí misma.

Mele formula sus condiciones suficientes, observa Holton, en términos de un individuo autoengañado respecto de la formación de una creencia particular. Sugiere, por el contrario, que debería tratarse al autoengaño como algo que ocurre con respecto a un tema o asunto. Propone un nuevo conjunto de condiciones suficientes, en el cual la primera condición de Mele es sustituida por: “el cuerpo de creencias de S acerca de  $\alpha$  contiene creencias erróneas acerca del sí mismo”<sup>12</sup>. Esta condición, no obstante, debe ser refinada. Antes de proceder a ese refinamiento, Holton recuerda el significado de “errores acerca del sí mismo”. Se trata de un error en un tipo de creencias que, si fuera el caso de que uno estuviera en lo correcto, contarían como autoconocimiento. Analizar qué es lo que se quiere decir con “autoconocimiento” es una tarea compleja y que Holton no pretende emprender; frente a ello, adopta una caracterización vaga, consistente en afirmar que se trata de conocimiento de uno mismo que está de algún modo íntimamente conectado con la propia psique. Si bien no avanza mucho más en este análisis, con la esperanza de que lo dicho sea suficiente para poder proceder al refinamiento de la condición modificada, sí añade que el

---

<sup>12</sup> Holton, R., o.c., p. 63.

autoconocimiento es un conocimiento *de se*: yo no poseería autoconocimiento en virtud de saber todo acerca del funcionamiento psíquico de un individuo que, desconocido para mí, resultara ser yo.

Dada la redefinición de la primera condición, el autor se pregunta ahora si las cuatro condiciones son suficientes para el autoengaño. Presenta aquí un ejemplo modificado del racista Jean Marie. Al presentar el caso, Holton supuso que Jean Marie no cumplía con sus propios estándares al momento de formar sus creencias racistas. Se encontraba en un error sobre sí mismo y, por lo tanto, autoengañado. Ahora bien, es posible imaginar un racista que cumpla completamente con sus propios y extravagantes estándares epistémicos. Supóngase, continúa Holton, que esta persona sinceramente piensa que debería creer solamente en lo que sus amigos le dicen, amigos que, a su vez, resultan ser racistas. Más aun, supóngase que no tiene ninguna razón ulterior para pensar esto; es su fundamento epistémico último [*epistemic bedrock*]. Tal persona, observa Holton, no necesita estar autoengañada; no estaría cometiendo el mismo error que Jean Marie: no pensará que está cumpliendo con estándares que, en realidad, no satisface. Sin embargo, Holton considera que en otro respecto sí estaría cometiendo un error acerca de sí mismo: creerá que sus propios estándares lo conducirán a creencias verdaderas, con lo que cumplirá con las cuatro condiciones prescriptas. Peor aún, prosigue, este es solo un caso de un problema muy general, ya que parecería que toda creencia falsa dará lugar a una creencia falsa sobre sí mismo, esto es, la creencia de que aquello en lo que uno cree es verdadero. En consecuencia, es posible encontrar en muchos casos que hay creencias que no son sobre el sí mismo, en el sentido que él está buscando y que, por ello, no constituyen casos de autoengaño; las creencias involucradas no son tales que si el agente estuviera en lo correcto, constituirían autoconocimiento.

El problema parece estar en que el error surge, en primera instancia, en una creencia que no es acerca de uno mismo. Por ejemplo, el racista que posee los estándares epistémicos extravagantes se forma la creencia general de que sus estándares lo conducirán a la verdad, y luego los aplica a sus propias prácticas. Si el error acerca de uno mismo da lugar al autoengaño, entonces el error debe comenzar por uno mismo. Por consiguiente, el autor distingue, introduciendo una nueva terminología, dos clases de errores acerca del sí mismo: aquellos que resultan de la aplicación a uno mismo de una creencia errónea que no es acerca de uno mismo, y aquellos que no (llama a los segundos *errores fundamentalmente de se*). Sugiere denominar a los casos que ocurren cuando las personas

cometen errores acerca de sí mismas en modos que, si hubieran estado en lo correcto, habrían constituido autoconocimiento, *errores de autoconocimiento*. Reuniendo lo anterior con las consideraciones previas, reformula la condición primera y necesaria de la siguiente manera: “1\*. El cuerpo de creencias de S acerca de  $\alpha$  contiene errores de autoconocimiento fundamentalmente *de se*”<sup>13</sup>.

### III

El análisis de Holton no es suficiente, a nuestro modo de ver, para probar que el autoengaño requiere como condición necesaria la existencia de creencias falsas acerca del sí mismo. Hay dos aspectos del autoengaño descuidados por el autor en este análisis, aspectos que redundan en que su posición requiera ser notoriamente debilitada, al punto de resultar trivialmente verdadera: el carácter temporal en el autoengaño y la distinción entre creencias de primer y segundo orden. Para la argumentación tendiente a mostrar que la condición propuesta por Holton no es necesaria no hará falta discutir el refinamiento que, al respecto de tal condición, propone el autor; esto es, el argumento se aplicará a cualquier caso de creencias referidas al sí mismo y, *a fortiori*, a los errores de autoconocimiento fundamentalmente *de se*.

Como hemos adelantado en la primera sección, Holton señala correctamente (aunque no sea ese su objetivo principal) que el autoengaño involucra un cuerpo de creencias, y no una única creencia falsa. Habitualmente se predica de alguien que está autoengañado respecto de *una* creencia, pero esta manera de describir el fenómeno podría generar la impresión engañosa de que quien se encuentra en estado de autoengaño solo posee una creencia falsa, es decir, la que es producto de procesos cognitivo/afectivo/motivacionales (para el caso de los modelos no intencionalistas) o resultado de la *intención* de adquirir tal creencia falsa (en el caso de los modelos intencionalistas). Como ha observado Oksenberg-Rorty<sup>14</sup>, el autoengaño, al igual que ocurre con otras formas de engaño, se multiplica. Apelando al ejemplo de una médica oncóloga que se autoengaña al interpretar lo que son señales absolutamente inequívocas de que padece un cáncer terminal, Oksenberg-Rorty señala que no solo se engaña acerca de su cáncer: para mantener esta falsedad tiene también que engañarse

---

<sup>13</sup> Holton, R., o.c., p. 65.

<sup>14</sup> Cf. Oksenberg-Rorty, A., “The Deceptive Self: Liars, Layers and Lairs”, en: McLaughlin, B. y A. Oksenberg-Rorty (eds.), *Perspectives on Self-Deception*, Berkeley: University of California Press, 1988, pp. 11-28.

a sí misma acerca de sus maniobras autoengañosas, el significado de sus enfoques, desvíos y negaciones. Si el autoengaño, observa esta autora, involucra algo más que estar equivocado o en conflicto, entonces parecería requerir algunas actitudes de segundo orden, en sí mismas también autoengañosas; algún reconocimiento de los conflictos entre las creencias y algunas estrategias *ad hoc* para reconciliar estos conflictos. En otras palabras, es posible distinguir entre diversos niveles de creencias, ordenados según se trate de creencias principales sobre el tema del autoengaño (por ejemplo, la creencia falsa acerca de nuestro buen estado de salud), o creencias derivadas de la necesidad de dar coherencia y sustentar las creencias principales sobre el tema del autoengaño (por ejemplo, la creencia en que los médicos que nos han diagnosticado una enfermedad grave no son profesionales competentes).

El hecho de que el autoengaño involucre un conjunto de creencias, por consiguiente, no elimina la posibilidad de establecer una organización o jerarquía dentro de ese conjunto. Parece útil, a los fines de este análisis, hacer una distinción entre creencias centrales y creencias periféricas en los casos de autoengaño. Sin negar que el autoengaño implique la existencia de un sistema complejo de creencias, parece razonable suponer que no todas las creencias tienen la misma jerarquía. La que llamaremos “creencia central” es la creencia falsa fundamental del autoengaño (por ejemplo, que nuestra pareja nos es infiel, cuando no lo es; que tendremos un desempeño excelente en un examen que debemos rendir, cuando todas las evidencias apuntan en sentido contrario; que nuestra salud es buena, cuando tenemos pruebas sólidas de que padecemos una enfermedad grave)<sup>15</sup>. Las creencias periféricas serán el conjunto de creencias desarrolladas alrededor de la creencia central, consistentes y mutuamente sustentantes (por ejemplo, que el hecho de que nuestra pareja alguna vez se haya demorado más de lo normal en volver del trabajo es prueba sólida de engaño; que el hecho de que hayamos descuidado la preparación de ciertos temas importantes para el examen no reduce nuestras probabilidades

---

<sup>15</sup> El examen de fenómenos emparentados con el autoengaño, como ciertos estados patológicos, contribuye a sustentar la idea de una creencia o tema dominante en la percepción distorsionada de la realidad. Por ejemplo, en el *Manual Diagnóstico y Estadístico de los Trastornos Mentales*, en su cuarta edición, se hace referencia, respecto de los subtipos de trastorno delirante, al “tema central de la idea delirante”. De este modo, quien padece este trastorno desarrolla un sistema de creencias delirantes alrededor de un tema central, por ejemplo, la creencia de ser objeto de engaño por parte de la pareja (en el tipo celotípico), de ser perseguido, espiado, engañado o víctima de una conspiración (en el tipo persecutorio), o de poseer algún talento o intuición extraordinario y no reconocido, o de haber hecho un descubrimiento importante (en el tipo de grandiosidad).

de aprobarlo; que el hecho de que los médicos que evaluaron el deterioro de nuestro estado de salud sean reconocidos especialistas en su área de trabajo no es significativo a la hora de aceptar el pronóstico)<sup>16</sup>.

Si se admite la distinción anterior, es posible clasificar en términos generales a los casos de autoengaño en dos tipos: aquellos en los cuales el objeto de la creencia central es algún estado del mundo (por ejemplo, el ya mencionado caso de la creencia errónea en la infidelidad de nuestra pareja), y aquellos en los cuales el objeto de la creencia central es el sí mismo (por ejemplo, creer falsamente que somos superiores en todo respecto a nuestros colegas de la universidad)<sup>17</sup>. Es obvio que en los casos de autoengaño en los que la creencia central versa sobre nosotros mismos podremos decir sin temor a equivocarnos que el estado de autoengaño incluye creencias erróneas sobre nosotros mismos e implica una falla en el autoconocimiento. Sin embargo, esto es mucho más dudoso en el caso de que la creencia central verse sobre algún objeto del mundo externo.

La admisión de que el autoengaño implica la existencia de un sistema de creencias erróneas estructuradas alrededor de una creencia central no significa que las actitudes de segundo orden, el reconocimiento de conflictos entre las creencias y las estrategias *ad hoc* para reconciliar estos últimos (en términos de Oksenberg-Rorty) impliquen la existencia en acto de todas las creencias correspondientes derivadas de esos procesos. Es aquí en donde el carácter temporal del autoengaño juega un rol decisivo. Muchas creencias erróneas surgirán como consecuencia de diversos procesos subsecuentes a la adquisición de la creencia central falsa: entre otros factores posibles, la confrontación de tal creencia por parte de personas cercanas, el hallazgo de evidencia que la contradice, o una personalidad tendiente a la duda y al autocuestionamiento.

Supongamos una vez más el caso del individuo celoso convencido falsamente de la infidelidad de su pareja. Más allá de la dificultad en determinar las motivaciones de semejante estado, parece aceptable la idea de que los mismos mecanismos cognitivos<sup>18</sup> que hacen posibles los casos de autoengaño “positi-

---

<sup>16</sup> Esto no excluye la posibilidad de que el autoengañado posea sospechas de que su creencia central es falsa.

<sup>17</sup> No creo que la distinción entre el sí mismo y el mundo sea una distinción carente de problemas; sin embargo, dado que Holton la emplea implícitamente en su planteo, me considero eximido de dar mayores precisiones sobre ella.

<sup>18</sup> Es de destacar que los mecanismos cognitivos que sesgan los procesos de inferencia y conducen a conclusiones incorrectas escapan a nuestra captación conciente. Es materia de debate, especialmente en los análisis más recientes acerca de la responsabilidad moral por el autoengaño (al respecto, *cf.* Levy, N., “Self Deception and Moral Responsibility”, en:

vos” conducen al sujeto a adoptar una creencia falsa contraria a la evidencia disponible. Ahora bien ¿posee el sujeto efectivamente creencias erróneas sobre sí mismo, o estas creencias son un producto ramificado y no necesario de procesos cognitivos posteriores? Imaginemos un proceso en el cual la creencia del celoso es puesta en tela de juicio por un interlocutor externo. En particular, el interlocutor cuestiona la manera en que el celoso ha evaluado la evidencia a su disposición. Ante el cuestionamiento, el celoso responde afirmando su ecuanimidad y equilibrio tanto en la búsqueda como en la interpretación de la evidencia. Cabe preguntarse: ¿poseía el celoso creencias falsas relativas a sí mismo, o estas creencias surgen en el proceso –proceso por otra parte no siempre presente–. Salvo que se recurra a la dudosa (y posiblemente *ad hoc*) noción de que las creencias preexistían en forma inconciente en la mente del sujeto celoso, y son simplemente explicitadas ante el cuestionamiento; parece más plausible la respuesta que sugiere que tales creencias no son más que una ramificación doxástica del proceso de autoengaño. Se puede admitir, también, que es verdad que un proceso de autocuestionamiento puede conducir a la formación de creencias falsas sobre sí mismo, pero este proceso no es necesario. Entiendo, al menos, que la carga de la prueba recae en quien diga que los procesos de autoengaño *siempre* implican el autocuestionamiento o alguna otra clase de proceso generador de creencias de segundo orden y la consecuente formación de creencias falsas acerca de sí mismo.

Es posible afirmar, en síntesis, que es muy plausible la idea de que el autoengaño en muchos casos (quizás incluso la mayoría) involucra la existencia de creencias falsas acerca de uno mismo, de una falla en el autoconocimiento; no obstante, parece poco plausible la idea de que *siempre* deben estar presentes estas creencias para poder hacer una atribución de autoengaño.

---

*Ratio* (New series), XVII (2004), pp. 294-311), la medida en la que podemos ser concientes de la actuación de estos mecanismos y, en tal medida, reducir su influencia sobre nuestras inferencias. Si esto fuese posible, posiblemente tendríamos, al menos en parte, algunas creencias falsas de la clase requerida por la reformulación de Holton; esto es, el agente puede llegar a creer que ha neutralizado las posibles fuentes de sesgo sobre sus inferencias y ha arribado a conclusiones correctas. Sin embargo, parece haber acuerdo en que el número y el carácter contraintuitivo de estos mecanismos hace sumamente difícil que su actuación pueda ser detectada y neutralizada de manera sistemática; más aun, es posible pensar que, en ciertos casos de autoengaño, su actuación no solo no será detectada, sino que el agente no se planteará ningún interrogante acerca de la legitimidad de sus propias creencias y la forma que las ha obtenido. De esta forma, no existirá la falla en el autoconocimiento requerida por Holton. Para una caracterización de un buen número de los mecanismos que sesgan las inferencias (cf. Thagard, P., “Critical thinking and informal logic: Neuropsychologic perspectives”, en: *Informal Logic*, XXXI, 3 (2011), pp. 152-170).

Una última objeción a la posición de Holton es que, a nuestro modo de ver, no ha probado de manera convincente que el contraejemplo que ofrece respecto de las condiciones suficientes de Mele (el caso de Catherine) no constituya un ejemplo de autoengaño. Sostener, como hace Holton, que Catherine simplemente rehúsa revisar cierta evidencia de la que dispone podría ser considerado como una descripción engañosa del caso. Si Catherine se abstuviera de revisar la evidencia pertinente respecto del nuevo trabajo, pero suspendiera a la vez el juicio sobre este, entonces sería claro que no se trataría de un caso de autoengaño; esto es, no se cumpliría la condición relativa a la falsedad de la creencia sustentada por el agente. Sin embargo, Catherine piensa, erróneamente, que el nuevo trabajo es claramente mejor que el actual. ¿Por qué Catherine rechazaría revisar cierta evidencia a su disposición si no existieran motivos poderosos para hacerlo? De este modo, podría ser igualmente plausible sospechar que Catherine *teme* que la proposición “el nuevo trabajo será mucho peor que el actual” sea verdadera y esto, sumado a su deseo de que la proposición opuesta (“el nuevo trabajo será mucho mejor que el actual”) sea verdadera, hace que rehúse revisar la evidencia pertinente. La reticencia a revisar la evidencia, por consiguiente, no es un dato que diferencie nítidamente el caso de Catherine de otros casos característicos de autoengaño.

Puede ser ilustrativo comparar el caso de Catherine con otro posible ejemplo de autoengaño intencional. Al respecto, podemos examinar en primer término el caso de Ike, propuesto por Mele<sup>19</sup> en su discusión relativa a las variedades de autoengaño. Ike, un olvidadizo bromista que posee habilidades para la imitación de la escritura manuscrita de otras personas, ha engañado intencionalmente a sus amigos escribiendo entradas falsas en los diarios de estos. Ahora bien, Ike ha decidido engañarse a sí mismo introduciendo una falsa entrada en su propio diario. Conciente de su falta de memoria, escribe bajo la fecha del día “estuve particularmente brillante en la clase de hoy”, contando con que, eventualmente, olvidará que tal afirmación es falsa. Unas semanas después, cuando revisa su diario, lee la oración y adquiere la creencia de que su desempeño fue brillante en la clase de ese día.

Mele observa que si Ike engaña a otros por medio de la introducción de notas falsas en sus diarios, no parecería haber nada que impida sostener justificadamente que se engaña intencionalmente a sí mismo: intenta causar en sí mismo la creencia de que *p*, sabiendo al mismo tiempo que *p* es falsa, y ejecuta esa

---

<sup>19</sup> Cf. Mele, *Self-deception Unmasked*.

intención sin ningún inconveniente, causando en sí mismo la creencia de que *p*. La definición estándar de engaño, observa Mele, es causar la creencia en aquello que es falso, y la causación en sí mismo por parte de Ike de una creencia falsa no es menos intencional que la causación en sus amigos de creencias erróneas.

El caso de Ike, observa Mele, nos impresiona como marcadamente disímil a los ejemplos de la “variedad de jardín” de autoengaño [*garden-variety examples*] (por ejemplo, el caso de la mujer que cree falsamente que su hijo no usa drogas frente a sólidas pruebas de lo contrario). La razón de esta impresión es que la diferencia más obvia entre el caso de Ike y los ejemplos de autoengaño de la “variedad de jardín” es la naturaleza directamente intencional del primero. Ike dispone de manera conciente engañarse a sí mismo, e intencional y concientemente ejecuta su plan para lograrlo, mientras que en los casos de autoengaño común las personas se comportan de manera muy diferente.

Si Mele está en lo correcto y el caso de Ike es inequívocamente un ejemplo de autoengaño, entonces parecería razonable sospechar que, *a fortiori*, el de Catherine también lo es. En ambos casos existe una intención deliberada con respecto a la falsa creencia: así como Ike intencionalmente induce en sí mismo una creencia falsa que desea adoptar, Catherine omite intencionalmente examinar con cuidado la evidencia con la que dispone, con el fin de evitar correr el riesgo de que la creencia predilecta, que a la postre resulta ser errónea, sea puesta en tela de juicio. Ejemplos como el de Ike, no obstante, no son unánimemente considerados casos de autoengaño. Audi<sup>20</sup>, por ejemplo, sugiere distinguir entre autoengaño y engaño autocausado, y considera que este último caso se ajusta a las características del segundo fenómeno y no a las del primero.

Sin embargo, no hace falta profundizar en la discusión acerca de si casos como el de Ike ejemplifican adecuadamente procesos de autoengaño intencional o alguna otra clase de fenómeno de irracionalidad motivada. Pese a las semejanzas señaladas, tal caso presenta también diferencias importantes con el de Catherine, diferencias que posibilitarían quizás clasificarlo en una categoría distinta de autoengaño intencional. Una diferencia fundamental reside en que, mientras Ike *sabe* que la creencia que intenta inducir en sí mismo es falsa, Catherine desconoce en realidad el valor de verdad de la afirmación falsa “el nuevo trabajo es claramente mejor que el actual”; simplemente opta por creer que lo es, sobre la base de lo que parece una revisión superficial de la evidencia con la que dispone y evita intencionalmente el examen cuidadoso de tal evidencia.

---

<sup>20</sup> Audi, R., “Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele”, en: *Behavioral and Brain Sciences*, 20 (1997), p. 104.

Siguiendo a Perring<sup>21</sup>, es posible considerar que el caso de Catherine se ajusta a un segundo tipo de autoengaño intencional no examinado por Mele. Esta variante se produce por medio de evitar e ignorar intencionalmente la contemplación de evidencia que se posee y que conduce a una conclusión perturbadora, teniendo a la vez pleno conocimiento de que se está otorgando prioridad a la tranquilidad mental actual antes que a la búsqueda de la verdad. Perring sugiere, a fin de ilustrar esta posibilidad, la siguiente adición a un caso propuesto por Mele<sup>22</sup>. Sam ha estado casado dos veces y ambos matrimonios han terminado en amargos divorcios. Ha estado muy ocupado con su trabajo, y está agotado cuando llega a su casa. Cuando su amigo cercano le dice que Sally ha sido vista con el señor Jones, Sam se estremece de temor y ansiedad; reconoce el significado potencial de la información. Tiene profundas convicciones religiosas y no estaría con Sally en caso de saber que ella le es infiel. Se dice a sí mismo “no voy a pensar acerca de esto”; enciende el televisor, bebe una cerveza y no vuelve a pensar acerca de la evidencia esa tarde, y de hecho tampoco lo hace las semanas posteriores, hasta que se encuentra directamente enfrentado con ella.

Perring observa que la evasión y la negación son formas comunes de autoengaño. En este caso, Sam intenta mantener la calma y evitar el dolor de pensar acerca de otro divorcio. Su autoengaño intencional no es autoderrotante [*self-defeating*], pero eso no se debe a que la intención de evitar evaluar la evidencia que posee esté oculta para él. Tal intención es muy explícita, y en el momento en que toma la decisión de evitar evaluar la evidencia que tiene, observa Perring, la búsqueda de la verdad queda en un segundo lugar respecto de su necesidad de mantener un equilibrio psicológico. El autoengaño es exitoso, observa el autor, dado que él no reflexiona sobre esto posteriormente, sino que se sumerge en otras actividades. Son la evasión y el involucramiento inicial lo que constituye el autoengaño intencional.

---

<sup>21</sup> Perring, C., “Direct, fully intentional self-deception is also real”, en: *Behavioral and Brain Sciences*, 20 (1997), pp. 123-124.

<sup>22</sup> “Sam ha creído durante muchos años que su esposa, Sally, nunca tendría un affaire. En el pasado, su evidencia para esa creencia era muy buena. Sally obviamente lo adoraba, nunca demostró interés sexual en otro hombre, condenó las actividades sexuales extramatrimoniales, era segura y feliz con su familia, y así siguiendo. Sin embargo, las cosas comenzaron recientemente a cambiar de modo significativo. Sally llega ahora tarde de su trabajo un promedio de dos noches por semana, frecuentemente encuentra excusas para abandonar la casa sola después de la cena, y Sam ha sido informado por un amigo cercano que Sally ha sido vista en compañía de un cierto señor Jones en un teatro y en un salón local. Sin embargo, Sam continúa creyendo que Sally nunca tendría un affaire. Desafortunadamente, está equivocado. La relación con Jones no es en absoluto platónica” (Mele, A., “Real Self Deception”, en: *Behavioral and Brain Sciences*, 20 (1997), pp. 91-136, pp. 95-96).

Las similitudes entre la reformulación del caso de Sam hecha por Perring y el caso de Catherine son impactantes: en primer lugar, en ambos casos existe una creencia central falsa; en segundo lugar, en ambos se antepone concientemente el logro de un objetivo no epistémico (conservar la tranquilidad de espíritu en un caso, mantener la esperanza en el otro) a la búsqueda de la verdad; por último, y de manera destacada, ambos agentes se abstienen intencionalmente de revisar evidencia que potencialmente refutaría la creencia predilecta. Parece difícil evitar la conclusión de que estamos ante una variante de autoengaño intencional diferente a la examinada por Mele, interpretación que compite con ventaja, a nuestro modo de ver, con la sugerencia de Holton de que se trata de un mero caso de “ceguera deliberada”.

Podría objetarse, en síntesis, que el ejemplo propuesto pertenece a un subtipo de autoengaño deliberado y conciente, pero autoengaño al fin. Si esta interpretación alternativa es plausible, entonces el argumento de Holton resulta aun menos convincente; los casos examinados no muestran sólidamente la necesidad de sustituir la primera condición sugerida por Mele. La carga de la prueba recae sobre Holton: él debe construir un caso mucho más sólido que el de Catherine para demostrar que las cuatro condiciones de Mele no son suficientes para el autoengaño.

Hemos afirmado que Holton está en lo correcto al señalar que el autoengaño involucra un cuerpo de creencias falsas, y no meramente una creencia singular, aunque discrepamos en lo que respecta a la composición y estructura que debe tener tal cuerpo de creencias. Ahora bien, cabe preguntarse si, eliminado como necesario el requisito de que tal cuerpo de creencias debe incluir creencias erróneas sobre el sí mismo, aquella observación es lo suficientemente fuerte como para justificar la sustitución de la condición de la creencia falsa propuesta por Mele y otros autores por una condición que haga referencia a un cuerpo de creencias. Dadas las notorias dificultades en la caracterización del fenómeno del autoengaño, la argumentación favorable a esta posibilidad excedería ampliamente los alcances de este artículo. Si, no obstante, se desea avanzar en esta dirección, quizás lo siguiente podría constituir una sugerencia útil para una condición modificada: “S adquiere o mantiene un cuerpo de creencias falsas, entre las cuales se encuentra la creencia falsa *p*, que se encuentra en la cima de un sistema jerárquico y coherente de creencias subordinadas o derivadas de ella”. Este cuerpo de creencias podrá, pero esta no es una condición necesaria, incluir creencias falsas sobre el sí mismo, esto es, puede o no implicar una falla en el autoconocimiento.