Autoridad y autoengaño

Marc Jiménez Rolland
Departamento de Filosofía
Universidad Autónoma de Aguascalientes
marcji2121@yahoo.com

Estuve revisando concienzudamente la literatura de Davidson en torno al problema del autoengaño y decidí presentarle un escrito a mi asesor sobre esta temática. Creo firmemente que he comprendido bien el asunto y me parece que el argumento, tal como lo he esquematizado, resulta bastante plausible. Le entrego el trabajo a Jorge y, tras revisarlo con detenimiento, me dice que he perdido la pista y que lo que presento ni siquiera se asemeja al argumento davidsoniano [supongamos que no me explica por qué]. Considero que mi asesor es una persona bastante competente en estos asuntos —no conozco a una persona con mayor conocimiento de ellos— y estoy convencido de que no está jugándome una broma, puesto que la situación no lo amerita. Tengo, pues, muy buenas razones para creer que mi asesor tiene razón y que yo estoy equivocado. Entregar este ensayo no es mi único curso de acción posible; después de todo, podría presentar un ensayo sobre alguno de los otros temas que he estado revisando. Aun así, decido presentar mi trabajo y sigo convencido de que mi presentación hace justicia al argumen-

to de Davidson. Si me lo preguntaran diría que mi asesor tiene razón en su juicio sobre el texto; pero diría también que considero que el escrito es adecuado. Creo que el trabajo es pertinente y también creo que no lo es.

Esto parece un auténtico caso de autoengaño. ¿Cómo se explica esta situación?

***

En lo que sigue me ocuparé de esbozar una caracterización general del autoengaño desde algunas de las principales perspectivas en la materia, señalando los problemas centrales que involucra y haciendo énfasis en la línea intencionalista. Luego analizaré la relación que el autoengaño guarda con la autoridad de la primera persona y qué imagen resulta de una conjunción teórica de ambos problemas. Finalmente intentaré mostrar que la situación planteada líneas arriba es en realidad sólo hipotética.

El autoengaño es un problema filosófico vasto que ha preocupado principalmente a la ética en lo relativo a la responsabilidad, la buena vida y la autenticidad de un individuo; sin embargo, también ha suscitado interés en los campos de la filosofía de la mente, la epistemología y la filosofía de la psicología. En el presente escrito trataré únicamente lo que concierne a estas últimas áreas, dejando de lado los asuntos morales.

Virtualmente todos los aspectos del autoengaño son materia de controversia en la discusión filosófica actual, incluyendo su definición y los casos paradigmáticos. Deweese-Boyd (2006), pretendiendo ser neutral frente al tema, señala que, en general, puede decirse que el autoengaño consiste en la adquisición y mantenimiento de una creencia (o, al me-

nos, la admisión de esa creencia) frente a fuerte evidencia de lo opuesto, «motivada por deseos o emociones» que favorecen la adquisición y retención de esa creencia. Sin embargo, la última parte de su definición, como veremos más adelante, favorece cierto enfoque del problema.

El interés filosófico en el autoengaño radica, entre otras cosas, en que supone una suerte de desviación epistemológica: un individuo racional adquiere y mantiene una creencia que se opone al total de la evidencia que posee (en ocasiones por motivaciones extrarracionales: deseos, temores, inquietudes, etc.). En el ejemplo con que inicio este escrito, yo podría seguir mi decisión de presentar el texto y pretender salir airoso —con bastante evidencia en contra de esto último— debido a que quiero evitarme la tarea de volver a redactar otro ensayo. Esto puede constituir una motivación válida para querer presentar mi ensayo (puede ser también la causa para que lo haga), pero no es una razón —o al menos no es una buena razón[1].

El autoengaño (*self-deception*), junto al pensamiento desiderativo (*wishful thinking*), la debilidad de la voluntad (*weakness of the will*) o *akrasia*, y el mal razonamiento (*bad reasoning*), es considerado por Davidson una forma de irracionalidad (ver 1997, 218; también 1982; 1985; 1986).

La forma tradicional en que se ha caracterizado el autoengaño, según comenta Deweese-Boyd (2006), es modelándolo sobre la base del engaño interpersonal: un sujeto *A intencionalmente* hace creer a un sujeto *B* la proposición *p*, sabiendo o creyendo que su contraria (~*p*) es verdadera. Se

[1] «If one would be happier, prouder, more relaxed, less fraught if one had a certain belief, that is a reason, putting other considerations aside, to have the belief. But such a reason is not, in itself, a reason to suppose the desirable belief is true» (Davidson 1997; 216).

considera, pues, una conducta intencional. Esta caracterización permite distinguir al autoengaño (y al engaño interpersonal) del mero error.

Sin embargo, Alfred Mele (1997) señala que la concepción intencionalista del autoengaño, modelada sobre la base del engaño interpersonal, da lugar a dos paradojas: la paradoja «estática» y la «dinámica». La primera no es otra que la paradoja de Moore: afirmar una proposición y decir enseguida que uno no se la cree; pero esto, como señala Mele (1997), requiere que el autoengañado se encuentre en un estado imposible de la mente: afirmar la contradicción. La segunda paradoja se refiere a la forma misma en que se modela el autoengaño intencional: hacer algo intencionalmente implica hacerlo a sabiendas; de modo que cuando uno se engaña a sí mismo lo hace sabiendo que quiere engañarse; pero esto supondría que uno sabe de antemano que lo que se pretende creer es falso.

Debido a estas paradojas, algunos autores como Mele (1997) y Funkhouser (2006) han optado por una aproximación no-intencionalista al autoengaño. Asumen que lo único que interviene en casos auténticos de autoengaño son motivaciones (deseos, aspiraciones, temores, inquietudes, etc.); de modo que su estrategia es deflacionista: el autoengaño es, para ellos, sólo una forma de error motivado por uno mismo.

El ejemplo planteado al inicio resulta ilustrativo, pues puede representar varias formas de irracionalidad. El caso podría ejemplificar una instancia de debilidad de la voluntad: mi mejor razonamiento sería, si considerara que el trabajo es inadecuado siguiendo la opinión de mi asesor, presentar algo sobre otra temática. Si no hiciera esto actuaría, no sólo contra el bienintencionado consejo de mi asesor, sino contra mi mejor juicio. Sin embargo, hay un aspecto que hace que

el ejemplo no sea un mero caso de debilidad de la voluntad: hasta el último momento creo que el trabajo es pertinente.

Por otra parte, como ya se ha señalado, conlleva una forma de pensamiento desiderativo: es mi deseo de no tener que adaptar un escrito sobre otro tema lo que motiva mi decisión de presentar ése y no otro texto. Hasta aquí un teórico motivacionalista estaría de acuerdo en que lo anterior representa un caso auténtico de autoengaño; más aún, diría que lo anterior **es todo lo que se requiere** para que el ejemplo constituya un caso de autoengaño. Sin embargo, pierde de vista un elemento importante: hace del autoengaño algo no-racional, suprimiendo su carácter irracional, que parece ser uno de sus elementos esenciales; «*for the irrational is not merely the non-rational, which lies outside the ambit of the rational; irrationality is a failure within the house of reason*» (Davidson 1982; 169). En el ejemplo inicial —y en muchos casos de autoengaño— se presentan en el sujeto dos creencias contrarias; y esto es algo de lo que el teórico motivacionalista no da cuenta, puesto que su estrategia está diseñada precisamente para evitar estas paradojas.

Pero renunciar a explicar este detalle es, en parte, renunciar a explicar el autoengaño, ya que es precisamente esto lo que lo distingue de otras formas de irracionalidad. Para dar cuenta de ello, algunos partidarios de la línea intencionalista han señalado que son divisiones temporales en la mente las que aíslan al autoengañado de su intención de engañarse (ver Bermúdez 2000). Esto explicaría lo que sucede con el protagonista de la película *Memento*, que intencionalmente decide ocultar evidencias de sus actos (descubre que fue engañado respecto a la muerte de su esposa y asesina a quien le mintió) para conservar su proyecto de vida (seguir buscando al asesino de su esposa); pero el personaje se encuentra en una situación muy peculiar: no puede generar recuerdos a

corto plazo, olvida en minutos lo que acaba de hacer. No es esto lo que ocurre en todos (ni siquiera en la mayoría de) los casos de autoengaño.

La propuesta de Davidson tiene la peculiaridad de incluir los aspectos anteriores e incluso explicar algunos más (puesto que es bastante compleja en detalles técnicos, pecaré de simplificación); además, goza de otra ventaja: se genera en el marco [del proyecto] de una teoría general del pensamiento, el significado y la acción (ver 1980).

Davidson parte del carácter normativo de la racionalidad; esto es, ser racional implica, entre otras cosas, seguir las normas básicas de la lógica, el principio de la evidencia total para el razonamiento inductivo y el principio de continenecia. Estos principios son compartidos por todas las criaturas que tienen actitudes proposicionales o actúan intencionalmente, puesto que es una condición para tener pensamientos, juicios e intenciones que los estándares básicos de la racionalidad tengan aplicación[2] (ver Davidson 1985; 195). No tiene sentido preguntar, de alguien con actitudes proposicionales, si es generalmente racional, pues esto es lo que se requiere para tenerlas; asimismo, los agentes no pueden **decidir** si aceptan o no los atributos fundamentales de la racionalidad: si están en condiciones de decidir en absoluto, poseen tales atributos.

Para explicar el rol causal de la motivación en el autoengaño, Davidson apunta, en primer lugar, algunas consideraciones sobre la explicación. Aunque los sucesos mentales en

[2] Debido precisamente a su concepción del vínculo entre pensamiento, lenguaje y acción aunado al carácter normativo de la racionalidad, Davidson señaló en alguna ocasión: «Lo que he repetido a menudo es que los niños, antes de tener gobierno sobre su habla, o los animales, no tienen de ninguna manera pensamientos. Esto parece molestar a personas que tienen mascotas, ¡y supongo que debe molestar a gente que tiene niños!» (Davidson 1992; 78).

alguna forma **causan** otros sucesos (sean mentales, como las creencias, o físicos, como las acciones), para la explicación recurrimos más bien a **razones**. Todo lo que el agente hace lo hace por una razón, a cuya luz la acción correspondiente es razonable (*reasonable*); las razones que el agente da de su acción ofrecen la intención con la que actuó, y por tanto la explican. Esa explicación debe existir si lo que hace una persona debe contar como una acción (ver Davidson 1982; 173). Sin embargo, hay una manera en que los procesos mentales **causan**, pero no **racionalizan** la acción: en los casos de irracionalidad motivada hay una causa mental (deseo, temor, etc.) que no es una razón.

Asimismo, Davidson reconoce que en algunos casos de autoengaño la división temporal es un factor explicativo importante (ver 1985; 198); pero ni estos ni los anteriores son los casos de autoengaño más interesantes, sino aquellos en los que la inconsistencia intrínseca es un factor determinante: cuando coexisten dos creencias contrarias. Es preciso reconocer el carácter holístico de lo mental; esto es, el significado de una oración, el contenido de una creencia o deseo, no son elementos que puedan ser atados a ellas en aislamiento de sus compañeros (otras oraciones; otras creencias o deseos); de ahí que estos tipos de autoengaño resulten problemáticos.

En este punto se precisa hacer explícita una distinción: es posible creer proposiciones contrarias simultáneamente ($Cp \wedge C{\sim}p$), pero no es posible creer la contradicción [$C(p \wedge {\sim}p)$]. Lo que ocurre en casos de autoengaño es lo primero; así se responde a la paradoja estática. El problema que esto suscita es cómo pueden las creencias coexistir en un mismo sujeto sin dar lugar a la conjunción. Aquí Davidson reconoce una deuda teórica con Freud: un recurso explicativo que podría funcionar para esta situación es la partición de la mente (ver

1997; 217). Sin embargo, a diferencia de Freud, Davidson no se vale del carácter metafórico de esta afirmación: las partes de la mente están definidas en términos de funciones; en última instancia, en términos de los conceptos de razón y de causa. Si las creencias contrarias se presentan ha de ser en partes distintas de la mente, de modo que el agente pueda creer ambas proposiciones sin caer en la contradicción.

Por otra parte, al parecer la partición de la mente —como parte del fenómeno del autoengaño— entra en contradicción con la autoridad de la primera persona, pues ésta consiste en la presunción, esencial para la posibilidad de interpretación —y, por ende, para la comunicación misma—, de que el hablante sabe lo que cree cuando afirma que cree algo; si el hablante tiene creencias contradictorias, tal presunción parece desorientada. De modo que la concepción de Davidson del autoengaño parece entrar en conflicto con la autoridad que él mismo reconoce. Sin embargo, esto no tiene por qué ser así: la incompatibilidad es sólo de grado. Es evidente que cuando uno descubre en un tercero un caso auténtico de autoengaño, la autoridad de éste en el conocimiento y la adscrispción de sus estados mentales es puesta en entredicho, mas no desaparece. El hecho de que él tenga creencias contradictorias no le da a un tercero ninguna primacía en el conocimiento o la adscripción de estados mentales; la autoridad se recupera eventualmente, cuando la persona se percata del autoengaño y se desengaña.

Cabe señalar que apostar por la introspección infalible —como lo hace Shoemaker (ver Zimmerman 2002)—, sobre la base de consideraciones motivacionalistas, puede resultar también una inferencia explicativa válida para casos de autoengaño, pero resulta mucho más restrictiva al considerar ciertas instancias de irracionalidad. Así mismo, una teoría

que no pudiera explicar la irracionalidad tampoco podría, en último término, explicar la autocrítica y la mejora en el propio pensamiento (consideradas elementos importantes de la racionalidad).

BIBLIOGRAFÍA

Bermúdez, J. (2000). Self-deception, intentions and contradictory beliefs. *Analisys*, 60 (4). [Online resource: http://www.artsci.wustl.edu/~pnp/Papers/bermudez2000d.pdf]

Davidson, Donald. (1980). A unified theory of thought, meaning, and action. First published as: Towards a unified theory of thought, meaning, and action. In *Grazer Philosopische Studie*n, (11), 1-12. Reprinted in Davidson (2004), 151-166.

_____. (1982). Paradoxes of irrationality. First published in *Philosophical essays on Freud.* R. Wollhein & J. Hopkins (eds.). UK: Cambridge University Press. 189-305. Reprinted in Davidson (2004), 169-187.

_____. (1985). Incoherence and irrationality. Fist published in *Dialectica, (39),* 345-354. Reprinted in Davidson (2004), 189-198.

_____. (1986). Deception and division. First published in *The multiple self.* J. Elster (ed.). UK: Cambridge University Press. 79-92. Reprinted in Davidson (2004), 199-212. [Traducción al español: Engaño y división. En Davidson, Donald. (1992). *Mente, mundo y acción*. *Claves para una interpretación*. Trad. de Carlos Moya Espí. Barcelona: Paidós, UAB | ICE. 99-117.]

_____. (1992). Respuesta a Mauricio Beuchot. *Quinto simposio internacional de filosofía*. Vol. I. Enrique Villanueva (comp.). México: UNAM.

_____. (1997). Who is fooled? First published in *Self-deception and paradoxes of rationality.* J.-P. Dupuy (ed.). Stanford, CA: CSLI, Stanford University Press. 15-27. Reprinted in Davidson (2004), 213-230.

_____. (2004). *Problems of rationality*. New York, NY: Oxford University Press.

Deweese-Boyd, Ian. (2006). Self-deception. *Stanford Encyclopedia of Philosophy.* USA: Metaphysics Research Lab, CSLI, Stanford University. [Online resource: http://plato.stanford.edu/entries/self-deception/]

Funkhouser, Eric. (2005). *Do the self-deceived get what they want?* Fayetteville, AR: University of Arkansas. [Online paper: http://comp.uark.edu/~efunkho/selfdeception.pdf]

Mele, Alfred. (1997). Real self-deception. *Behavioral and Brain Sciences*. [Online manuscript: http://www.bbsonline.org/documents/a/00/00/05/19/bbs00000519-00/bbs.mele.html]

Zimmerman, Aaron Zachary. (2002). *Infallible introspection*. NEH Summer Seminar—University of California. [Online paper: http://consc.net/neh/papers/zimmerman.htm]

RESUMEN

En el presente artículo, el autor se ocupa de esbozar una caracterización general del autoengaño desde algunas de las principales perspectivas en la materia, señalando los problemas centrales que involucra y haciendo énfasis en la línea intencionalista. Luego analiza la relación que el autoengaño guarda con la autoridad de la primera persona y qué imagen resulta de una conjunción teórica de ambos problemas.

Palabras clave: Davidson; autoengaño; autoridad; racionalidad.

ABSTRACT

In this article, the author is concerned with outlining a general characterization of self-deception from some of the top prospects in the field, noting the central problems involved and emphasizing the intentionalist line. Then he analyzes the relationship that keeps self-deception with the authority of the first person and what image results as a theoretical combination of both.

Keywords: Davidson; self-deception; authority; rationality.