

**LA NATURALEZA DEL CONOCIMIENTO DE
NUESTRO MUNDO MENTAL:
ACERCA DE UNA TRILOGÍA QUE IMPORTA:
AUTOCONOCIMIENTO, DELIBERACIÓN Y ENGAÑO, DE
NORA STIGOL, ALICIA GIANELLA Y DIEGO LAWLER**

**The Nature of the Knowledge of our Mental World:
On *Una trilogía que importa: Autoconocimiento,
deliberación y engaño*, by Nora Stigol, Alicia Gianella,
and Diego Lawler**

GUSTAVO FERNÁNDEZ ACEVEDO ^a
<https://orcid.org/0000-0003-0542-7380>
facevedo@mdp.edu.ar

^a Facultades de Humanidades y Psicología, Universidad Nacional de Mar del Plata, Argentina.

Resumen

El libro que es objeto de esta nota crítica expone un amplio conjunto de problemas y perspectivas teóricas sobre el fenómeno del autoconocimiento y otros estrechamente vinculados con él, especialmente, el del autoengaño. Examinaremos aquí tres cuestiones cercanamente vinculadas entre sí: la relación entre autoconocimiento y autoengaño, el rol de la racionalidad en el autoconocimiento, la medida en que las creencias sobre nosotros mismos son susceptibles a las influencias volitivas y, por último, la posibilidad de que coexistan distintos modelos explicativos.

Palabras clave: Autoconocimiento; Autoengaño; Explicación; Racionalidad; Voluntarismo doxástico.

Abstract

The book under review presents a comprehensive set of problems and theoretical perspectives on self-knowledge and related issues, particularly self-deception. We will examine three closely interconnected questions here: the relationship between self-knowledge and self-deception, the role of rationality in self-knowledge, the extent to which beliefs about ourselves are susceptible to volitional influences, and, finally, the possibility that different explanatory models may coexist.

Key-words: Self-knowledge; Self-deception; Explanation; Rationality; Doxastic Voluntarism.

1.

La naturaleza del autoconocimiento, cuya importancia desde la Antigüedad resulta condensada en la máxima “conócete a ti mismo”, atribuida al Oráculo de Delfos, constituye una fuente inagotable de complejos interrogantes filosóficos. La vigencia de tales interrogantes se mantiene hasta nuestros días, como bien muestra el volumen colectivo del que nos ocupamos en esta nota, producto de una serie de reuniones desarrolladas en la Sociedad Argentina de Análisis Filosófico. Entendido como tema central para la filosofía de la mente, el concepto de autoconocimiento resulta crucial en dos sentidos: para la posibilidad de ser un agente epistémico (lo que requiere una familiaridad con los propios pensamientos, sensaciones, experiencias perceptuales y acciones) y para alcanzar cierta sabiduría y llevar una buena vida, intuiciones que, señalan los autores, permean nuestra vida cotidiana. Ahora bien, la relevancia y antigüedad del problema no les impide observar (señalamiento que estará presente a lo largo del libro) que la expresión “autoconocimiento” mantiene un grado no menor de ambigüedad, esto es, que admite distintas acepciones que suponen diferentes maneras de abordar el fenómeno. Esta ambigüedad los conduce a establecer con claridad los límites de su análisis. La noción del Yo, así como el intento de establecer o adoptar una teoría de la mente y su ontología, y los aspectos sociales, políticos y éticos derivados de la manera de concebir el autoconocimiento quedan fuera de su examen, sin que por esto se los estime de menor importancia. El foco de su interés estará puesto, entonces, en la pregunta acerca de cómo accedemos al conocimiento de nuestra propia vida mental, nuestras creencias, nuestros deseos y emociones. Sobre esta base, presentan distintos problemas y perspectivas teóricas sobre el autoconocimiento y otros fenómenos más o menos estrechamente vinculados con él, como los de autoignorancia, autoexamen, autoconciencia, autoría, autoridad de la primera persona y, especialmente, autoengaño.

El enfoque adoptado para el examen de los problemas relacionados con el autoconocimiento, dado el recorte descrito, está determinado por los debates filosóficos contemporáneos, en particular, por la perspectiva de la filosofía analítica de corte anglosajón. Esto no impide que los autores adviertan que recientemente diversas disciplinas científicas, como la psicología, la sociología, la antropología y la biología se han ocupado de su estudio y efectuado contribuciones de interés para nuestra comprensión del fenómeno.

Es de destacar la variedad y complejidad de los problemas y temas relacionados con el autoconocimiento. Este conjunto de problemas incluye el de la agencia, el de la naturaleza de las creencias en particular y de

las actitudes proposicionales en general, el problema mente-cuerpo, el del estatus de la psicología de sentido común e incluso otros que, aun cuando no aparezcan de modo explícito, pueden conectarse con él de una manera relativamente directa en nuestra opinión. Estos serían el problema de la identidad personal (basta notar, al respecto, que uno de los criterios examinados para abordar este problema fue el de la continuidad psicológica que implicaría, en principio, la posibilidad del autoconocimiento),¹ y el de la naturaleza del libre albedrío (al menos para ciertos modelos, como el denominado “modelo de la autoría”, según el cual somos seres racionales capaces de autodeterminarnos cognitiva y prácticamente).

Dado que el volumen está compuesto por siete capítulos que difieren entre sí tanto en lo que respecta a los problemas planteados como a las perspectivas teóricas examinadas, parece conveniente dividir el texto subsiguiente en dos partes. En la segunda sección expondremos sintéticamente los problemas estudiados, junto con sus propuestas de solución y, en algunos casos, ciertas dificultades específicas de ellos. En la tercera sección, y a partir de la síntesis previa, desarrollaremos algunos nexos entre las cuestiones teóricas esbozadas en los distintos capítulos que, a nuestro modo de ver, justifican un mayor examen.

2.

En el capítulo 1, A. Gianella sienta las bases para el estudio del problema del autoconocimiento, proponiendo tres estrategias para abordarlo: a) la identificación de varios aspectos (teóricos, fácticos, lingüísticos, históricos y evaluativos), b) el análisis de los conceptos intervinientes, esto es, el conocimiento y el yo, y c) el examen de las nociones vinculadas con la de autoconocimiento, entre otras, las de autoengaño, autoexamen, y autoconciencia.

En los siguientes tres capítulos, N. Stigol examina distintas perspectivas teóricas sobre el autoconocimiento, partiendo de las más tradicionales, como el cartesianismo, hasta algunas de las más recientes y refractarias a esa tradición. En el curso de este examen reitera la advertencia hecha en la Introducción respecto de la ambigüedad del término “autoconocimiento”, y enuncia la necesidad de efectuar un recorte respecto de sus alcances.

En el capítulo 2 estudia las posiciones que consideran al autoconocimiento como un logro epistémico, esto es, aquellas posiciones que, justamente, se denominan epistémicas o cognitivas, y para las cuales el autoconocimiento comprende un proceso por el cual el sujeto llega a conocer

¹ Cfr., por ejemplo, Rodríguez González (2009).

sus propios estados mentales. Estas perspectivas son el cartesianismo, el modelo perceptual (o del sentido interno) y el modelo inferencial. El cartesianismo sienta las bases para un gran número de debates posteriores a partir de la postulación de una serie de rasgos que caracterizarían el autoconocimiento, entre los que se encuentran la infalibilidad, la exhaustividad, la transparencia, la autoridad de la primera persona y el acceso privilegiado. La autora señala que resulta cuestionable el supuesto del carácter incorregible de nuestro autoconocimiento, y también que fenómenos como el autoengaño, el pensamiento desiderativo y la *akrasia* resultan incompatibles con la infalibilidad. Los modelos perceptual e inferencial, si bien comparten la perspectiva epistémica respecto del autoconocimiento y a este como un logro cognitivo, reflejan apartamientos considerables del modelo cartesiano. El primero explica el conocimiento de nuestra propia mente en términos análogos al modo en que se explica nuestro conocimiento perceptual del mundo externo. El segundo, por su parte, tanto en sus versiones conductistas (Ryle, Dennett), como en sus variantes más modernas y sofisticadas (Cassam), rechaza aspectos centrales del modelo cartesiano, como la asimetría de acceso y la autoridad de la primera persona.

En el capítulo 3 examina las posiciones que niegan que el autoconocimiento sea un tipo de conocimiento y, en consecuencia, que sea un logro cognitivo. Entre ellas están el constitutivismo, la estrategia de la autoría y el expresivismo. Para el constitutivismo, el autoconocimiento no resulta de una conexión cognitiva entre un estado mental de primer orden y uno de segundo orden, sino que se trata de una conexión *constitutiva* entre ellos. El tener conciencia del estado mental en el que me encuentro no implica otro acto mental que acompaña al primero, sino que ambos son uno y el mismo acto; el acceso a mis estados mentales no es una propiedad extrínseca y relacional. El constitutivismo difiere del cartesianismo tanto en lo que respecta al camino elegido para defender sus tesis principales (lo hace en términos ontológicos, no gnoseológicos), como en el rechazo a los rasgos que el segundo atribuye al autoconocimiento: incorregibilidad, infalibilidad, omnisciencia, etcétera.

En el capítulo 4 examina la perspectiva expresivista desarrollada por Dorit Bar-On, que constituye un intento de conciliación entre las estrategias epistémicas y no epistémicas. En términos generales, el expresivismo es el punto de vista que defiende la idea de que la función de las emisiones en los diversos ámbitos del discurso no es la de describir o representar hechos del mundo, sino la de expresar nuestros sentimientos, actitudes y emociones, así como otros estados mentales. La posición de Bar-On se distancia de las formas menos sofisticadas del expresivismo (“expresivismo simple”), razón por la cual se la ha denominado “neoexpresivismo”. Todas

las variantes del expresivismo se proponen explicar el autoconocimiento partiendo del análisis de las “declaraciones” [*avowals*] mediante las cuales un sujeto se autoatribuye estados mentales en primera persona de tiempo presente, como “siento dolor”. Las declaraciones no describen ni informan acerca de “estados internos” del sujeto; no representarían conocimiento ni involucrarían juicios o creencias acerca de algún tipo de entidades. Ambas versiones coinciden en pensar las declaraciones por analogía con las emisiones no lingüísticas a las que reemplazan, tales como gritos, gemidos, muecas y otros; no están basadas en evidencia alguna y reemplazan las expresiones naturales.

En los capítulos 5 y 6, D. Lawler examina la denominada “estrategia de la autoría”. En el capítulo 5 caracteriza esta estrategia a partir de la distinción de dos posiciones respecto del rasgo de la transparencia. Mientras que algunas posiciones filosóficas se muestran muy impresionadas por esta característica, y la consideran la puerta de entrada para la identificación de los aspectos clave del autoconocimiento, otras rechazan completamente este supuesto. El enfoque de la autoría, debido a R. Moran, que se sitúa dentro del primer grupo, sostiene que lo que está en juego en el autoconocimiento tiene menos que ver con que un agente goce de un acceso especial a su mundo mental que con el hecho de que lo que piensa *depende* de él. Por ejemplo, si una persona declara que cree que su pareja no le ha dicho la verdad, no está informando sobre un contenido de su vida mental, sino que está manifestando su resolución sobre un asunto, resolución que expresa la autoridad de su razón sobre su propia vida mental. Este enfoque sugiere entonces un giro práctico al problema del autoconocimiento y la autoridad de la primera persona: tal autoridad se debe a que, en cierta medida, es un asunto que depende de ella. Ahora bien, esta noción de dependencia o autoría no es fácil de esclarecer; la noción que interesa al autor se relaciona con aquello que es un hecho distintivo de nuestra condición de seres racionales, esto es, que contamos con la capacidad de autodeterminarnos cognitiva y prácticamente. Somos seres reflexivos, y esta capacidad, en la forma de autorreflexión, está involucrada en la determinación racional de nuestros pensamientos.

En el capítulo 6, Lawler examina la naturaleza práctica del modelo de la autoría. Este modelo no sólo involucra una autoconcepción del sujeto que conoce, sino también una determinada relación con el objeto de este conocimiento, su vida mental. El sujeto no puede concebirse a sí mismo como sometido única y exclusivamente al orden causal y a la receptividad pasiva; los contenidos mentales son en alguna medida resultado de su actividad de conocerlos (nos conocemos al transformarnos). Esta posición pretende mantener al mismo tiempo la idea de que la autoridad de la primera perso-

na expresa un logro cognitivo genuino. Ahora bien, si existe tal logro, debe aceptarse que el sujeto que conoce responde a una realidad independiente del hecho de conocerla. No obstante, el sujeto debe ejercitar su condición de *agente* sobre los contenidos de su condición psicológica. Si se siguiera esta vía argumentativa, prosigue, el conocimiento en primera persona que se podría obtener de los contenidos proposicionales de una condición psicológica particular pertenecería a lo que Anscombe denominó “conocimiento práctico”. Mediante, justamente, la aplicación del modelo de ese conocimiento elaborado por esta filósofa, Lawler examina qué clase de conocimiento es el autoconocimiento involucrado en la estrategia de la autoría.

En el capítulo 7, el más extenso del volumen, A. Gianella estudia el fenómeno del autoengaño. Identifica para su examen los mismos cinco aspectos que propuso para el estudio del autoconocimiento, y señala los aportes de distintas disciplinas para su comprensión. Presenta las paradojas tradicionalmente reconocidas en los estudios sobre el fenómeno (las paradojas estática y dinámica o estratégica), su naturaleza de fenómeno irracional, cuestiona el rol privilegiado que la creencia, en tanto actitud proposicional, ha tenido en su estudio, y diferencia tipos o clases de autoengaño y su relación con otros fenómenos de “irracionalidad motivada”, como la *akrasia* y el pensamiento desiderativo.² Este extenso análisis conduce a la autora a un balance final que incluye tanto conclusiones como conjeturas. Dos cuestiones centrales sobresalen respecto del autoengaño, ambas de tipo explicativo: cómo y por qué se produce. No obstante, también resalta la importancia de una cuestión no explicativa, sino normativa, esto es, la relativa a si es beneficioso autoengañarse. El balance, que incluye respuestas a las cuestiones señaladas, resalta las siguientes conclusiones: a) en el autoengaño hay intereses vitales que prevalecen sobre el conocimiento. Un sujeto se autoengaña debido a que la posesión de la creencia le provee de ventajas en su vida; b) operan sobre nosotros soluciones de compromiso, en una suerte de economía interna. El autoengaño es una forma de solución de conflictos que parece irracional si se lo evalúa exclusivamente bajo los estándares de la racionalidad epistémica; c) se debe buscar alternativas a las descripciones del autoengaño que provienen de la concepción clásica, y sustituir la idea de contradicción lógica por la noción de disociación psicológica; d) la noción estándar de creencia no es la más satisfactoria para estudiar el autoengaño; puede haber alternancia de creencias contradictorias, y no hay que descartar alternativas, como pensamientos, registros

² Cabe señalar que hubiera sido sumamente pertinente el examen de las formas “negativas” de autoengaño, esto es, aquellos casos en los cuales la persona autoengañada adquiere o sostiene creencias que considera indeseables y que, a la vez, no están adecuadamente justificadas por la evidencia de que dispone.

no conscientes, “saber” en sentido débil y otras; e) nuestra mente no tiene partes, sino subestructuras que son plásticas y dinámicas; es innecesario sostener divisiones “robustas” para explicar el autoengaño; f) pagamos un alto costo por autoengañarnos: el abuso de este recurso puede empobrecernos como personas, ya que proporciona una respuesta fácil y rápida para lidiar con los desafíos de la vida que, a la vez, nos disminuye cognitivamente y emocionalmente; g) existe cierta racionalidad en el autoengaño; constituye una meta pendiente lograr la determinación de criterios de racionalidad que permitan evaluar qué formas del autoengaño son racionales (en caso de que las hubiera) y cuáles no.

3.

Como señalamos en la primera sección, la estructura del libro no admite la identificación de una o más tesis que lo unifiquen, aunque sí lo hacen, por supuesto, la perspectiva filosófica adoptada y el rigor en el tratamiento de los problemas examinados. Sí creemos que hubiera sido enriquecedora una mayor búsqueda de nexos conceptuales entre los diferentes problemas y perspectivas estudiadas. Esta ausencia se advierte en mayor medida, a nuestro modo de ver, en los posibles nexos entre autoconocimiento y autoengaño, máxime si se tiene en cuenta que, como señala Gianella en el capítulo 1, “dada la complejidad y la relevancia de este fenómeno, dedicamos un capítulo para analizarlo más detalladamente” (p. 30). Sobre esta base, identificaré tres puntos teóricos que conectan autoconocimiento y autoengaño: el rol de la racionalidad, la medida en que las creencias sobre nosotros mismos son susceptibles a las influencias volitivas y, por último, la posible oposición entre autoconocimiento y autoengaño. Todos estos puntos, como veremos, están interconectados, y tienen un carácter más bien propositivo como posibles líneas de trabajo a seguir. Para finalizar, examinaremos una cuestión de alcance mayor, referente a la medida en que los modelos explicativos del autoconocimiento pueden ser compatibles o incluso articulables entre sí.

a. Autoconocimiento, autoengaño y voluntarismo doxástico

Como vimos, en el capítulo 5 Lawler examina minuciosamente el denominado “modelo de la autoría” con respecto al autoconocimiento. No reiteraremos su caracterización general, pero sí resulta necesario para nuestros propósitos introducir la distinción que presenta el autor entre dos modelos concebidos para explicar intuiciones en apariencia contradictorias acerca del proceso deliberativo implicado en la condición de transparencia:

el “modelo de la cristalización” y el “modelo activo” (MA, en lo sucesivo). De acuerdo con el primero, si bien la deliberación respecto de si p acontece o no es un caso de agencia, la adquisición de la creencia correspondiente no lo es. De acuerdo con el segundo, la actividad de juzgar si p ocurre (o no) es parte del proceso deliberativo; consecuentemente, formar una creencia es también un caso del ejercicio de la agencia racional. Pese a que estos modelos representan intuiciones diferentes, Lawler sostiene que pueden articularse de un modo que permite elucidar dimensiones fundamentales del fenómeno de la transparencia.

Ahora bien, tal articulación requiere evitar un peligro conceptual que subyace al MA. Al examinarlo, Lawler plantea el serio desafío que representaría para su plausibilidad filosófica el que esta posición implique el voluntarismo respecto de la creencia. Esto es, si las acciones relativas a la decisión de que p es el caso fueran consideradas como un caso directo de agencia, entonces el suceso consistente en adoptar p estaría sujeto a la voluntad del agente, con lo cual la creencia podría ser adquirida por consideraciones que no constituirían razones adecuadas en su favor. Esta posibilidad, que implicaría la admisión de una posición filosófica extremadamente difícil de sostener, esto es, el voluntarismo doxástico (Williams, 1973),³ debe ser eludida si se desea defender la plausibilidad del MA.

Lawler elabora, respecto de este punto, una serie de intuiciones debidas a G. Watson (2004) que permitirían evitar los riesgos mencionados:

Para el MA, entonces, el proceso de formación de creencias comprendería dos clases de agencia; por una parte, una agencia intencional relacionada con la actividad de recolección de evidencias y su valoración; por otra parte, una agencia doxástica, a saber, una agencia no intencional, la cual tendría lugar durante el proceso de realizar juicios cognitivos, donde el sujeto es guiado por su competencia normativa, esto es, su sensibilidad justificadora —para decirlo de algún modo—. El ejercicio de esta clase de agencia no sería objeto del argumento de Williams, puesto que no muestra ninguna dependencia con deseos e intereses del sujeto, esto es, con ninguna clase de objeto de la voluntad del sujeto (p. 129).

Sin embargo, aun en caso de que la estrategia adoptada logre eludir el riesgo de implicar el voluntarismo directo respecto de la creencia, puede igualmente plantearse el interrogante de si consigue evitar todos los peligros derivados de esa posición. Este interrogante se deriva de la distinción

³ Existen autores, no obstante, que consideran insuficientes los argumentos de Williams (cfr., p. ej., Funkhouser, 2003).

entre el voluntarismo doxástico directo (que se ajusta a la noción inicialmente descripta) y el voluntarismo doxástico indirecto. Según esta última variante resultaría posible, en principio, inducir en nosotros mismos ciertas creencias de un modo indirecto. Esto se lograría por medio de la modificación planificada de nuestras conductas, alternativa que suele ser ejemplificada por el consejo de Pascal de actuar como si se fuese creyente, de modo de inculcar en nosotros mismos la creencia en Dios. El propio Williams (1973) quien, como Lawler correctamente señala, elaboró un sólido argumento en contra del voluntarismo doxástico directo, admitió que hay lugar para aplicar la decisión de creer a través de medios indirectos, aunque consideró que esta alternativa es profundamente irracional y que puede tener éxito solo por medio del autoengaño. El nexo entre voluntarismo doxástico indirecto y autoengaño ha sido debatido, discusión que ha incluido, contra Williams, la negación de la necesidad del segundo para el primero (cfr. Cook, 1987). Creemos, sin embargo, que no necesitamos pronunciarnos aquí respecto de este vínculo, aunque es plausible pensar que, si el voluntarismo doxástico indirecto implica un deterioro de la racionalidad, tal deterioro podría ser incluso mayor si el autoengaño se viera involucrado. Tal como ha sido descrito, el voluntarismo doxástico indirecto podría afectar el proceso deliberativo que conduce a la aceptación de la creencia tanto en el modelo de la cristalización como en el modelo activo. Si, como señala Lawler, “las creencias (u otras actitudes proposicionales) son el precipitado de la actividad racional del sujeto” (p. 131), la posibilidad de que la forma de voluntarismo descripta desempeñe un rol constituye una amenaza que, es plausible pensar, debería ser considerada seriamente.

b. Racionalidad

Como hemos visto, existe una conexión íntima entre el autoconocimiento y la racionalidad. Esta conexión se manifiesta de formas diferentes en los distintos modelos sobre el primero. Una de ellas tiene lugar en ciertos modelos constitutivistas, como el de Shoemaker (que es denominado, justamente, “racionalismo”), en el cual la racionalidad es concebida como condición de posibilidad del autoconocimiento. Por otra parte, las posiciones “agencialistas”, como las de Burge y Bilgrami, también consideran que nuestra naturaleza racional contribuye con la posibilidad de autoconocernos. Por último, también tiene un vínculo estrecho con el modelo de la autoría, como queda claramente expuesto en la presentación de Lawler. Ahora bien, más allá de la fortaleza del rol que se asigne a la racionalidad para la posibilidad de autoconocimiento, no hay duda de que existe una asociación positiva entre ambos.

Sin embargo, la contribución de la racionalidad de nuestras creencias para nuestro bienestar no es una tesis que carezca de aristas y controversias. Algunos filósofos, como Davidson (1985) y Williams (1973), han señalado que el autoengaño positivo puede eventualmente liberar a quien lo practica de pensamientos o realidades dolorosas de distinta clase. Esto es, la adquisición o mantenimiento de creencias no racionales (o cuya racionalidad es deficitaria) respecto de nosotros mismos podría eventualmente contribuir a nuestro bienestar o, en otros términos, a la posibilidad de lograr una buena vida. Esto abre la posibilidad de plantear lo siguiente: ¿podría ser instrumentalmente racional inducir en uno mismo creencias que se desvían de lo que estamos justificados en creer respecto de nosotros mismos debido a su presunta contribución a una buena vida? En otros términos, la pregunta es si puede existir, respecto del autoconocimiento, una suerte de “disociación” entre la racionalidad teórica y la racionalidad instrumental. La importancia de esta cuestión no es menor si, recordemos, se sostiene que el autoconocimiento es necesario para llevar una buena vida. Tal posibilidad ha sido examinada por algunos psicólogos (Taylor & Brown, 1988, 1994), y también por distintos filósofos (Badhwar, 2008; Van Leeuwen, 2009). Los primeros, en particular, examinaron los efectos de las denominadas “ilusiones positivas”, clase de fenómenos que se caracterizan por consistir en una visión injustificadamente favorable hacia nosotros mismos.⁴ Estas son concepciones positivas no realistas del sí mismo, exagerada percepción de control personal y optimismo no realista.⁵ Las ilusiones positivas podrían ser consideradas, en consecuencia, sesgos en el autoconocimiento.

Ahora bien, una cosa es admitir que, en algunos casos (cuyos límites sería razonable intentar establecer), la presencia de ilusiones positivas *puede* constituir una contribución al bienestar, y otra, muy distinta, es derivar de este juicio fáctico la prescripción de que resulta racional inducir en nosotros mismos esta clase de ilusiones. Si esto último fuera el caso, entonces estaríamos en presencia de lo que podría interpretarse como una inconsistencia entre racionalidad teórica y práctica. La primera indicaría un ajuste racional de nuestras creencias respecto de nosotros mismos (entre otras razones, para intentar lograr una buena vida), la segunda indicaría un desvío de esa norma (aunque no una contradicción completa), desvío

⁴ Las ilusiones positivas pueden ser consideradas variantes débiles del autoengaño, aunque su pertenencia a esa categoría no es totalmente clara.

⁵ Si bien los estudios de Taylor y sus colaboradores datan de algunas décadas atrás, la posibilidad de que las creencias ilusorias puedan desempeñar un papel positivo sigue siendo objeto de investigación hasta nuestros días tanto dentro de la filosofía como de la psicología (cfr. Makridakis & Moleskis, 2015; Schütz & Baumeister, 2017; Kaufmann et al., 2019; Ho, 2022).

justificado por sus presuntos efectos positivos hacia ese mismo fin. Esta tesis más fuerte es la que ha sido severamente criticada. Diversos autores, entre ellos Bhadwar (2008) y Van Leeuwen (2009) han presentado sólidos argumentos en su contra. Sin embargo, aun cuando se admitiera que la tesis más fuerte sea insostenible, podría ocurrir que la versión más débil tenga apoyo empírico y debilite el nexo entre la racionalidad de nuestras creencias respecto de nosotros mismos y, consecuentemente, la posibilidad de acercarse a una buena vida.

c. La relación entre autoengaño y autoconocimiento

La pregunta relativa a las relaciones entre autoengaño y autoconocimiento no parece admitir una respuesta única, lo que se debe, al menos en parte, a que las teorías sobre la naturaleza de cada uno de estos fenómenos son muchas y muy variadas, incluso incompatibles. Consecuentemente, cualquier respuesta resultará dependiente de la posición que se adopte respecto de cada uno de ellos. En el volumen objeto de esta nota tales relaciones aparecen respecto de dimensiones más globales, en algunos casos, y más locales, en otros, de los fenómenos examinados. Stigol, en varios pasajes, formula observaciones en esta segunda dirección. En algunos casos lo hace en el contexto del análisis de enfoques teóricos particulares, como la tradición cartesiana (respecto de esta, señala que solo si la infalibilidad e incorregibilidad son falsas es posible que fenómenos como el autoengaño se tornen inteligibles), y el expresivismo. En otros, en conexión con las explicaciones de autores como Bilgrami y Cassam respecto de la naturaleza del autoengaño. Lawler, a su vez, reflexiona respecto de la posibilidad de que la deliberación conducente a una conclusión respecto de qué creer pueda conducir al autoengaño. Esto puede ocurrir, por ejemplo, en el caso de que la reflexión autoconsciente esté dominada por una compulsión, situación en la cual puede desviarse y debilitar la autoridad que se asienta sobre ella.

Gianella es quien plantea de modo más directo el problema de los nexos entre autoconocimiento y autoengaño. En un pasaje del primer capítulo, señala lo siguiente:

En cuanto a la relación del autoengaño con el autoconocimiento, en algunos aspectos podemos pensar que son antagónicos, al menos parcialmente: si nos autoconocemos, no podríamos autoengañarnos; si nos autoengañamos, es porque no nos conocemos lo suficiente (p. 30).⁶

⁶ Con respecto a la segunda parte de la conjunción, también podría invertirse el orden causal: no nos conocemos lo suficiente debido a que nos autoengañamos.

Asimismo, en la p. 160 señala también Gianella que se ha afirmado que el autoengaño constituye tanto una falla en el autoconocimiento como un obstáculo en nuestro conocimiento del mundo externo.

Estas afirmaciones permiten formular distintos interrogantes vinculados entre sí. El supuesto de partida, esto es, que el autoengaño constituye un fenómeno opuesto (al menos en alguna medida) al autoconocimiento parece sumamente plausible. Como se ha observado reiteradamente, la posesión de creencias (o la actitud proposicional que se prefiera) verdaderas respecto de nosotros mismos y del mundo parece ser una condición básica no solo para la racionalidad tanto teórica como instrumental, sino también para nuestra propia subsistencia.⁷ Ahora bien, aceptado el supuesto, es posible plantear diversos interrogantes. ¿En qué medida el autoconocimiento es afectado por el autoengaño?; ¿afecta de modo diferencial a las distintas dimensiones del autoconocimiento?; ¿existe alguna clase de vínculo entre el autoengaño con respecto al mundo y el autoengaño respecto de nosotros mismos? Parece plausible, respecto del primer interrogante, la siguiente suposición: entre autoconocimiento y autoengaño no puede existir una oposición “global”, sino solo local. Esto es, el autoengaño podrá tener un mayor o menor lugar en nuestro sistema de creencias (si su producto es una creencia), pero parece plausible la suposición de que se trata de un fenómeno “localizado”, que tiene lugar en alguna parte de nuestros sistemas de creencias respecto de nosotros mismos, pero no se extiende a la totalidad. Dicho en otros términos, y apelando a una metáfora, el autoengaño constituiría un conjunto de islas en un océano de conocimiento de nosotros mismos. El antagonismo, como se sugiere en la cita, puede ser parcial, pero no total.

Ahora bien, el autoengaño podría constituir un obstáculo mucho mayor para el autoconocimiento en caso de que todo proceso de autoengaño, aun cuando la creencia producto del autoengaño refiera a fenómenos del mundo externo, involucre una falla en el autoconocimiento. Aquí se pueden plantear, como mínimo, dos posibilidades. Una de ellas, débil, de acuerdo con la cual una parte de los casos de autoengaño con respecto al mundo externo incluye el autoengaño con respecto a uno mismo; la segunda, fuerte, sostiene que *todos* los casos de autoengaño requieren, como condición necesaria, la presencia de creencias falsas acerca de uno mismo.⁸Cuál de las dos alternativas es correcta, es una cuestión controvertida. Ahora bien, aun

⁷ Cfr., entre otros, McKay & Dennett (2009).

⁸ Esta posición fue defendida por R. Holton (2000), quien sostuvo que la caracterización del autoengaño debe incluir necesariamente 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.

en caso de que la versión fuerte de la tesis sea incorrecta, de todos modos esto podría dejar lugar a la posibilidad de que un número indeterminado de casos de autoengaño *incluyan* creencias erróneas acerca del sí mismo; si esto fuese así, entonces el autoengaño constituiría un fenómeno opuesto en mayor medida al autoconocimiento que en el caso contrario.

Las cuestiones examinadas en a., b. y c. conducen a pensar, a nuestro modo de ver, en la necesidad de profundizar en las relaciones entre autoconocimiento y las diversas formas de apartamiento de la racionalidad, especialmente, el autoengaño, como estrategia para iluminar los aspectos más complejos de ambas clases de fenómenos. Si bien podría pensarse en una versión más fuerte de esta tesis, según la cual no habría manera de comprenderlos cabalmente examinándolos por separado, no hace falta adherir a ella: si lo planteado en los puntos a - c se considera plausible, esto sería suficiente para sostener la versión inicial de esta sugerencia.

d. Pluralismo explicativo y unicidad del autoconocimiento

Como último punto de esta nota queremos detenernos brevemente en una cuestión que, como señalamos, podría considerarse “metateórica” (si cabe el término en el ámbito filosófico), y que es la relativa a la posibilidad de coexistencia y complementariedad de distintos modelos explicativos del autoconocimiento. En el capítulo 3 Stigol señala lo siguiente:

Mi propósito al presentar enfoques explicativos distintos acerca del AC está guiado por la idea de que un único modelo teórico de explicación no es satisfactorio dada la rica multiplicidad y heterogeneidad de los estados mentales que constituyen nuestra vida psicológica. En mi opinión, está lejos de ser cierto que una única estrategia teórica que aspire a dar cuenta del fenómeno del AC pueda ser adecuada para explicar las diversas maneras como los seres humanos tenemos de acceder a la pluralidad y complejidad de nuestra vida mental. Creo, en cambio, que las distintas teorías acerca del AC no resultan ser teorías alternativas, sino complementarias (p. 66).

La sugerencia expresada en la cita podría ser considerada un ejemplo de lo que, en algunos contextos teóricos, sería denominado “pluralismo explicativo”. Si bien la encuentro *prima facie* plausible, la tesis genera diversos interrogantes cuyo tratamiento, entiendo, hubiera enriquecido el enfoque que se le da al problema estudiado en el libro. Dado que la sugerencia no es explorada en profundidad, constituye, a nuestro modo de ver, una línea adicional de análisis del volumen que estamos examinando.

La plausibilidad de la tesis deriva, en primer lugar, de la consideración de una dimensión ontológica: la multiplicidad y variedad de los fenómenos mentales que, en principio, pueden ser objeto de autoconocimiento, torna muy improbable la pretensión de que un único modelo explicativo pueda ser capaz de proporcionarnos una comprensión relativamente completa de su naturaleza. No obstante, es importante distinguir dos aspectos lógicamente independientes en la observación de Stigol. Por un lado, la propuesta de que podrían ser necesarios distintos modelos para explicar el fenómeno del autoconocimiento y, por el otro, la sugerencia de que tales modelos no resultan alternativos, sino complementarios. Esta última, creemos, es la que resulta más problemática. Por supuesto, este carácter problemático depende de en qué consista la posible complementariedad explicativa, cuestión que excede nuestros propósitos en este trabajo. Ahora bien, si “complementario” refiere a que los distintos modelos no son incompatibles entre sí, y conjuntamente ofrecen una visión completa de aquello que se estudia, la sugerencia parece ser, como dijimos, problemática. Las diferencias en los supuestos teóricos fundamentales de los distintos modelos generan dudas fundadas, a nuestro modo de ver, acerca de que tal complementariedad sea posible. ¿Pueden ser complementarios modelos explicativos que se basen, como supuesto fundamental, en la tesis de que el autoconocimiento es un logro cognitivo con aquellos que parten de la negación de tal supuesto? También resulta pertinente preguntar si pueden ser complementarias la teoría neoexpresivista y el modelo inferencial o, más en general, el expresivismo y las teorías que consideren al autoconocimiento un logro cognitivo y, en consecuencia, adopten una perspectiva epistémica del autoconocimiento. Si sus supuestos teóricos resultan inconciliables (como parece ser el caso, teniendo en cuenta las caracterizaciones presentadas), resulta difícil pensar en la compatibilidad de las explicaciones que ofrecamos de ellos. En cualquier caso, entendemos que la carga de la prueba recae sobre quien pretenda sostener alguna clase de complementariedad.

Si lo anterior resultara aceptable, sería posible conjeturar una consecuencia tal vez más radical respecto de la naturaleza del autoconocimiento. La variedad e incompatibilidad de los modelos explicativos podría considerarse un indicador de que lo que denominamos “autoconocimiento” es en realidad un conjunto heterogéneo de fenómenos abarcados laxamente bajo ese rótulo; dicho en otra manera, el término genera la ilusión de que estamos frente a un fenómeno unitario caracterizado por determinadas propiedades —cuáles son, esto es muy discutible—, cuando en realidad se trata de un conjunto o familia de fenómenos de límites difusos, tanto en lo que respecta a cada integrante de esa familia como en lo que concierne a los que no pertenecen a ella.

Más allá de cualquier reparo que hayamos presentado aquí (o cualquier otro que pudiera formularse), consideramos que se trata de un libro excelente para cualquier interesado en comprender los problemas relativos a la naturaleza del autoconocimiento y su conexión con una multiplicidad de interrogantes filosóficos.

Bibliografía

- Badhwar, N. K. (2008). Is realism really bad for you? A realistic response. *The Journal of Philosophy*, 105(2), 85-107. <https://doi.org/10.5840/jphil200810526>
- Cook, J. T. (1987). Deciding to believe without self-deception. *The Journal of Philosophy*, 84(8), 441-446. <https://doi.org/10.2307/2027001>
- Davidson, D. (1985). Engaño y división. En *Mente, mundo y acción*. Paidós.
- Funkhouser, E. (2003). Willing belief and the norm of truth. *Philosophical Studies*, 115, 179-195. <https://doi.org/10.1023/A:1025094823262>
- Ho, T. H. (2022). Positive illusion and the normativity of substantive and structural rationality. *Philosophical Explorations*, 26(3), 293-304. <https://doi.org/10.1080/13869795.2022.2160876>
- Holton, R. (2001). What is the role of the self in self-deception? *Proceedings of the Aristotelian Society*, 101(1), 53-69. <https://doi.org/10.1111/j.0066-7372.2003.00021.x>
- Kaufmann, M., Goetz, T., Lipnevich, A. A., & Pekrun, R. (2019). Do positive illusions of control foster happiness? *Emotion*, 19(6), 1014-1022. <https://doi.org/10.1037/emo0000499>
- Makridakis, S., & Moleskis, A. (2015). The costs and benefits of positive illusions. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00859>
- McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences*, 32(6), 493-561. <https://doi.org/10.1017/S0140525X09990975>.
- Rodríguez González, M. (2009). El problema de la identidad personal. En P. Chacón Fuertes, V. L. Guedán Pécker, J. A. Guerrero del Amo, J. Hermoso Durán, J. I. Morera de Guijarro & M. Rodríguez González, *Filosofía de la psicología* (pp. 279-301). Biblioteca Nueva.
- Schütz, A., & Baumeister, R. F. (2017). Positive illusions and the happy mind. En M. D. Robinson & M. Eid (Eds.), *The happy mind: Cognitive contributions to well-being* (pp. 177-193). Springer. https://doi.org/10.1007/978-3-319-58763-9_10
- Stigol, N., Gianella, A., & Lawler, D. (2024). *Una trilogía que importa: Autoconocimiento, deliberación y engaño*. EUDEBA.

- Taylor, S., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193-210. <https://psycnet.apa.org/doi/10.1037/0033-2909.103.2.193>
- Taylor, S., & J. Brown (1994). positive illusions and well-being revisited separating fact from fiction. *Psychological Bulletin*, 116(1), 21-27. <https://psycnet.apa.org/doi/10.1037/0033-2909.116.1.21>
- Van Leeuwen, D. S. N. (2009). Self-deception won't make you happy. *Social Theory and Practice*, 35(1), 107-132. <https://doi.org/10.5840/soctheorpract20093516>
- Watson, G. (2004). The work of the will. En *Agency and answerability: Selected essays* (pp. 123-158). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199272273.003.0006>
- Williams, B. (1973). Deciding to believe. En *Problems of the Self: Philosophical Papers 1956-1972* (pp. 136-151). Cambridge University Press. <https://doi.org/10.1017/CBO9780511621253.011>

Recibido el 27 de junio de 2025; aceptado el 3 de julio de 2025.