

<https://doi.org/10.21555/top.v760.3359>

Disfunción, desajuste y subdeterminación empírica

Dysfunction, Mismatch, and Empirical Underdetermination

Federico Burdman

Universidad Alberto Hurtado

Chile

fburdman@uahurtado.cl

<https://orcid.org/0000-0002-8506-9271>

Juan R. Loaiza

Universidad Alberto Hurtado

Chile

jloaiza@uahurtado.cl

<https://orcid.org/0000-0003-0570-3832>

Francisco Pereira

Universidad Alberto Hurtado

Chile

fpereira@uahurtado.cl

<https://orcid.org/0009-0008-9119-9487>

Recibido: 09 - 10 - 2024.

Aceptado: 15 - 09 - 2025.

Publicado en línea: 15 - 06 - 2026.

Cómo citar este artículo: Burdman, F., Loaiza, J. R. y Pereira, F. (2026). Disfunción, desajuste y subdeterminación empírica. *Tópicos, Revista de Filosofía*, 76, 219-253. <https://doi.org/10.21555/top.v760.3359>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Resumen

¿Son los trastornos psiquiátricos causados por el funcionamiento incorrecto de la maquinaria de la mente? ¿O pueden explicarse como el resultado de la falta de ajuste entre nuestros mecanismos psicológicos y un ambiente diferente a aquel en el cual estos sistemas evolucionaron? En este artículo, adoptamos elementos centrales de la concepción de la función propia desarrollada por Daniel Dennett para abordar una versión particular del debate entre teorías de la disfunción y teorías del desajuste, enfocándonos en casos de adicción a sustancias. Siguiendo la idea de Dennett según la cual no hay un hecho decisivo que determine la función propia de los sistemas biológicos, sostenemos que la decisión sobre la disfunción del sistema de recompensa en la adicción está subdeterminada por la evidencia. Nuestro argumento invita a reorientar la discusión, sugiriendo un nuevo foco en las consideraciones pragmáticas que pueden esgrimirse para elegir entre dos teorías empíricamente equivalentes.

Palabras clave: adicción; disfunción; desajuste; función propia; sistema de recompensa; subdeterminación; evolución; mecanismo; actitud de diseño; enfermedad.

Abstract

Are psychiatric disorders caused by a malfunction in the mind's mechanisms? Or can they be explained as the result of a mismatch between our psychological mechanisms and an environment different from the one in which these systems evolved? In this paper, we draw on central elements of Daniel Dennett's conception of proper function to address a specific version of the debate between theories of dysfunction and theories of mismatch, focusing on cases of substance addiction. Following Dennett's idea that there is no decisive fact determining the proper function of biological systems, we argue that the decision regarding the dysfunction of the reward system in addiction is underdetermined by the evidence. Our argument calls for a reorientation of the discussion, suggesting a new focus on the pragmatic considerations that can be invoked to choose between two empirically equivalent theories.

Keywords: addiction; dysfunction; misalignment; intrinsic function; reward system; underdetermination; evolution; mechanism; design stance; disease.

1. Introducción¹

En las sociedades modernas, una gran cantidad de personas sufre a causa de problemas relacionados con la salud mental. Muchos acuden a consulta con profesionales especialistas en salud mental y reciben, conforme a los protocolos usuales, un diagnóstico en términos de las categorías de la nosología psiquiátrica, junto con indicaciones para su tratamiento clínico o farmacológico. De acuerdo con una concepción usual en filosofía de la psiquiatría, tales condiciones de salud mental — al menos, las que merecen el nombre de enfermedad o trastorno— son el resultado de algún tipo de disfunción en los mecanismos subyacentes. La persona sufre y experimenta dificultades porque alguno de los sistemas que integran su vida psicológica está fallando (Boorse, 2014; Wakefield, 1992; Murphy, 2009, entre muchos otros). Llamaremos, de modo genérico, a las teorías que suscriben a este punto de vista *teorías de la disfunción*. Una concepción alternativa, inspirada en la teoría de la evolución por selección natural, propone que una porción significativa de las condiciones que ocupan a psiquiatras y psicólogos clínicos podría explicarse mejor por la observación de que las condiciones en las que vivimos actualmente son notablemente diferentes del entorno ancestral en que evolucionó nuestra especie y en el cual nuestros mecanismos psicológicos fueron moldeados. Aun cuando los mecanismos psicológicos actúen del modo previsto y presuntamente funcional, esto solo conduce confiablemente a resultados positivos cuando estos operan bajo las condiciones para las que fueron diseñados (Cosmides y Tooby, 1999; Murphy y Stich, 2000; Richters y Hinshaw, 1999, entre otros). De este modo, las *teorías del desajuste* (en inglés, *mismatch*) proponen explicar el malestar y el sufrimiento experimentados por los usuarios de servicios clínicos sin necesidad de postular la ocurrencia de una disfunción.

En este artículo, adoptamos elementos centrales de la concepción de la función propia desarrollada por Daniel Dennett para abordar una versión particular del debate entre teorías de la disfunción y teorías del desajuste, enfocándonos en casos de adicción a sustancias. Tomando como punto de partida una descripción comúnmente aceptada del modo en que funcionan los circuitos cerebrales asociados a la recompensa en

¹ Agradecemos a dos referís anónimos para *Tópicos* por sugerencias sumamente útiles para afinar nuestro argumento en diversos puntos.

casos de adicción, discutimos las razones a favor y en contra de sostener que la adicción involucra una falla en este sistema motivacional. Según nuestra reconstrucción, el punto central en que divergen las teorías de la disfunción y las teorías del desajuste concierne a un problema acerca de cómo caracterizar precisamente la función propia del sistema de recompensa: ¿es la función de dicho sistema rastrear y señalar eventos y situaciones que covarían con aumentos en la disponibilidad de dopamina, o es su función, en cambio, rastrear y señalar eventos y situaciones que aportan valor biológico al organismo, tales como la comida, el sexo, o la interacción social? Recurriendo a ideas de Dennett, sostenemos que ambas descripciones de la función del mecanismo son igualmente compatibles con todos los hechos relevantes acerca de su historia evolutiva. Esta subdeterminación conduce a una disolución productiva de la discusión entre teorías del desajuste y teorías de la disfunción. Nos permite, en primer lugar, ver que no hay ningún hecho decisivo al que podamos aludir para sostener que una de las descripciones es objetivamente más adecuada que la otra, de modo que la elección entre ambas no puede basarse en la acumulación de evidencias. Esto sugiere, en segundo lugar, la necesidad de reenfocar la discusión en una dirección diferente. Una vez que advertimos que no son los hechos acerca de la historia evolutiva lo que nos permitirá dirimir esta disputa, las opciones en juego pueden ser distinguidas a partir de consideraciones pragmáticas. Nuestra discusión señala, de este modo, la necesidad de explicitar y discutir cuáles han de ser los criterios pragmáticos que guíen la elección entre estas teorías.

Procederemos de la siguiente manera. En la sección 2, presentamos un breve estado del arte sobre los mecanismos neurofisiológicos involucrados en la adicción, con énfasis en el sistema de recompensa. Esto nos permitirá distinguir las teorías del desajuste y las teorías de la disfunción sobre el trasfondo de las hipótesis acerca de la función del sistema mencionado. En la sección 3, examinamos el argumento de Dennett según el cual no hay ningún hecho decisivo que permita elegir una única descripción, teoría o hipótesis correcta sobre la función propia de un sistema. A continuación, en la sección 4, aplicamos las ideas de Dennett al caso del sistema de recompensa y presentamos nuestro argumento principal. Sostenemos que de no haber un hecho decisivo que permita determinar la función propia del sistema de recompensa de manera unívoca, no hay entonces cuestiones de hecho que saldar en torno a si el sistema de recompensa es disfuncional, disolviendo la

discusión entre teorías del desajuste y teorías de la disfunción. Luego, en la sección 5, exploramos una objeción según la cual nuestro argumento se apresura al abandonar el proyecto de encontrar una descripción unívoca de la función propia del sistema de recompensa, lo que lleva a consecuencias inaceptables. Proponemos que no solo el argumento no lleva a tales consecuencias, sino que provee avenidas fructíferas y más enfocadas de discusión en torno a la adicción. Específicamente, concluimos con algunas reflexiones sobre el rol de los valores en la práctica científica y clínica en psiquiatría de la adicción, mostrando cómo nuestro argumento permite integrar consideraciones de valor en el debate sobre la función del sistema de recompensa.

2. ¿Hay una falla en el sistema de recompensa en la adicción?

Las teorías neurobiológicas de la adicción encuentran buena parte de su motivación en un hallazgo sumamente sugerente. Sabemos que hay un conjunto de sustancias —el alcohol, la nicotina, la cocaína, la marihuana, los opiáceos, entre otras— que producen, al ser consumidas, un repertorio variado de efectos sobre la experiencia, la cognición, la emoción y la conducta de quienes las consumen. A pesar de las diferencias significativas que hallamos en su composición química y en sus efectos, todas son potencialmente objetos de adicción. El hallazgo sumamente sugerente al que nos referimos es que todas estas sustancias producen un resultado neuroquímico similar a nivel cerebral: todas producen un aumento en la disponibilidad de dopamina (Nestler, 2005). Lo hacen a través de vías diferentes —algunas estimulan directamente la producción de dopamina, otras disminuyen la producción de sustancias que inhiben la producción de dopamina, otras reducen la recaptación de dopamina remanente en las juntas sinápticas, etc.— pero todas ellas conducen a un aumento en los niveles de dopamina disponible en sistemas clave del cerebro. La importancia de este hallazgo no es difícil de colegir: parece confirmar la presunción de que estas sustancias diversas, emparentadas por sus efectos sobre fenómenos de alto nivel, están unificadas también por un mecanismo común a nivel neuroquímico.

Esto sugiere, adicionalmente, una vía de explicación neurobiológica para su potencial adictivo. Esta se monta sobre una imagen bastante detallada que hemos llegado a formar en las últimas décadas de un sistema cerebral que tiene a la dopamina en el centro de la escena: el *sistema de recompensa*, una agrupación funcional que engloba a diferentes circuitos neuronales ubicados fundamentalmente en el mesencéfalo,

incluyendo porciones del área tegmental ventral, el núcleo accumbens, los ganglios basales, el tálamo y sus conexiones con diferentes circuitos de la corteza prefrontal medial, entre otras áreas (Grace *et al.*, 2007). Según la visión estándar, el sistema de recompensa es esencialmente un sistema motivacional, esto es, está involucrado en la regulación de la conducta orientada a metas, especialmente en la búsqueda y obtención de cosas tales como comida, agua, sexo e interacción social. Dado el evidente valor biológico de la obtención de tales cosas, es claro que estar equipados con un sistema que nos motive a conseguirlas repercutirá positivamente sobre la aptitud del organismo. En efecto, el sistema de recompensa es sumamente antiguo desde el punto de vista filogenético. Todos los mamíferos cuentan con un sistema de este tipo. Algo similar a él se encuentra incluso en especies de invertebrados (Nesse y Berridge, 1997, p. 64).

Aunque restan puntos de controversia respecto del modo exacto en que funciona este sistema, algunos de sus elementos centrales pueden describirse con confianza. Si describimos su operación en términos representacionales, el sistema funciona, en primer lugar, categorizando ciertos estímulos o eventos como recompensas.² El sistema representa aquello que se ha obtenido como valioso, y emite una señal que comunica esto al resto del cerebro, la cual corresponde típicamente en el plano fenomenológico con una experiencia placentera. Algunos estímulos pueden provocar esta respuesta de forma incondicionada, sin la mediación del aprendizaje. Se trata de lo que se conoce a veces como recompensas naturales o recompensas primarias (Butlin y Papineau, 2017).

Más allá de la señalización de la obtención de recompensas, el sistema tiene un papel fundamental a la hora de mover al organismo en la búsqueda de recompensas. En ello el aprendizaje tiene un papel fundamental, estableciendo asociaciones entre ciertos objetos y circunstancias y la obtención de recompensas. El mecanismo a través del

² Aquí y en lo que sigue adoptaremos tentativamente el uso de hablar del sistema de recompensa como un sistema representacional. Creemos que hay buenos motivos para ello, con independencia de cómo se especifiquen exactamente los contenidos representacionales involucrados. Esta elección no juega, sin embargo, un papel central en nuestro argumento, que podría funcionar igualmente bajo una concepción de la identificación de recompensas y/o del deseo en términos no representacionales.

cual esto sucede es bien conocido en psicología. Se trata del aprendizaje por refuerzo, un tipo robusto de aprendizaje mediado por recompensas y castigos. La obtención de cosas gratificantes puede constituirse entonces como un fin de la conducta orientada a metas. Esto supone que, en el lenguaje de la psicología folk, el sistema tiene un papel fundamental en la generación y sostenimiento de *deseos*. Se trata de deseos intrínsecos (*i. e.*, no instrumentales) por la obtención de recompensas, sean estos ocurrentes o disposicionales (Schroeder, 2004). La experiencia pasada en la obtención de recompensas media usualmente, a través del aprendizaje por refuerzo, la generación de disposiciones a desear ciertas cosas. En suma, el sistema está involucrado en la generación de deseos disposicionales, de largo plazo, por la obtención de recompensas, abarcando tanto a las recompensas primarias como a otras que el organismo aprendió a valorar a partir de su asociación con aquellas (Berridge, 2009; Dill y Holton, 2014).³

La dopamina tiene en todo ello un papel fundamental: el mecanismo neurocomputacional a través del cual el sistema señala las recompensas es una señal dopaminérgica. La dopamina está involucrada, en efecto, tanto en las experiencias placenteras como en la formación de deseos disposicionales.⁴ Esto permite explicar, por un lado, por qué el consumo de sustancias como las mencionadas está ampliamente difundido en

³ Cabe añadir aquí dos observaciones. En primer lugar, no pretendemos sugerir que todos los deseos sean generados de este modo por el sistema de recompensa. Algunos de nuestros deseos —los que Dill y Holton (2014) llaman *deseos cognitivos*— parecen estar mediados por juicios evaluativos que no juegan un papel en la operación del sistema de recompensa. En segundo lugar, de acuerdo con la hipótesis de Terry Robinson y Kent Berridge (2008), las vías neurales que subyacen a la señalización de recompensas presentes y a los aspectos motivacionales de los deseos disposicionales de largo plazo son suficientemente diferentes como para admitir la posibilidad de una disociación entre el disfrute (*liking*) y el deseo (*wanting*). Según Robinson y Berridge, esto juega un papel clave en la adicción, al permitir explicar cómo es posible que alguien pueda desear algo intensamente aun cuando no encuentre gratificante la experiencia de obtenerlo.

⁴ El modo exacto de describir el papel neurocomputacional de la dopamina en este proceso ha sido asunto de debate. Una hipótesis popular en la literatura científica propone que la señal dopaminérgica no marca propiamente la obtención de recompensas, sino la discrepancia entre el valor de la recompensa esperada y el efectivamente obtenido. De acuerdo con esta hipótesis, su papel neurocomputacional sería el de una *señal de error en la predicción* (Montague et

sociedades humanas de todo el mundo —se halla documentado ya en tiempos prehistóricos (Durrant *et al.*, 2009, p. 1049)— y por qué también muchas especies de animales no humanos disfrutaban del consumo de estas sustancias y las consumen cuando se les da la oportunidad. Más centralmente para nuestros fines, esto sugiere una visión relativamente concreta del mecanismo neurobiológico de la adicción. Cada vez que el organismo consume estas sustancias, se refuerza la asociación entre ellas y la obtención de recompensas. Como el impacto que estas sustancias tienen sobre la disponibilidad de dopamina es desproporcionado en relación con el que producen otras recompensas, el deseo disposicional por obtener más de ellas puede también ser desproporcionadamente fuerte en comparación con otros deseos. Dado que las drogas operan directamente sobre la señal neurocomputacional que media el aprendizaje por refuerzo a partir de la obtención de recompensas, el sistema es *engañado*, en cierto sentido, por ellas. Señaliza la obtención de algo de gran valor para el organismo, generando una disposición sumamente fuerte a buscar y obtener más de aquello, a pesar de que el objeto en cuestión tiene una contribución neutra, cuando no negativa, a la aptitud del organismo.

El resultado es que algunos sujetos desarrollan una disposición sumamente fuerte y persistente a buscar y obtener más de tales sustancias.⁵ Dado que se trata de deseos originados en este sistema motivacional básico (en oposición a deseos mediados por creencias y juicios evaluativos), estos tienden a persistir de un modo que es relativamente impermeable a consideraciones racionales y a los juicios de valor que el mismo agente forma a partir de la deliberación (Burdman, 2024; Holton y Berridge, 2013). El resultado puede ser extraordinariamente dañino. Para muchas personas, conduce a un espiral descendente que destruye todo lo que quieren y valoran, de modo que no hay mucho espacio para

al., 1996; Schultz, 2016). Cfr. Colombo (2014) para una revisión de la evidencia a favor y en contra de esta hipótesis.

⁵ Una nota importante es que solo una porción relativamente pequeña de los consumidores de drogas desarrolla el tipo de deseos problemáticos que encontramos en la adicción. Aunque sabemos que las personas varían en su susceptibilidad a la adicción en virtud de varios elementos, incluyendo factores genéticos (Verhulst *et al.*, 2015), no contamos en la actualidad con una explicación completamente satisfactoria de por qué el consumo lleva a la adicción en algunos casos y no en otros.

el debate respecto del carácter perjudicial de la adicción. Sin embargo, sí hay razones relevantes para preguntarse si tales resultados dañinos son, o no, el resultado de una *falla* en el funcionamiento del sistema de recompensa. En otras palabras: ¿está *funcionando incorrectamente* el sistema al señalar a las drogas como recompensas y al generar deseos disposicionales de largo plazo por consumirlas?

Esta pregunta plantea una cuestión importante para las teorías de la adicción, en la medida en que resume parte del debate acerca de la necesidad de postular la existencia de disfunción mecanística para explicar la conducta adictiva —una cuestión que tiene potencialmente implicaciones sobre el debate más amplio acerca del estatus de la adicción como enfermedad (Burdman, 2025)—. Más allá de eso, sin embargo, es una pregunta relevante por derecho propio en cuanto señala una cuestión conceptual importante acerca del modo en que entendemos el concepto de “función propia”.

A fin de abordar esta cuestión, haremos dos suposiciones. Por un lado, supondremos que algo similar a la descripción anterior es correcto y que no hay, en el contexto de la adicción, alteraciones estructurales o funcionales significativas en la operación de los mecanismos de recompensa. La discusión en la que nos enfocaremos gira en torno a la pregunta acerca de si es suficiente, para decir que el sistema funciona incorrectamente, que el sistema esté equivocado —en un sentido que intentaremos precisar— acerca del valor de las drogas para el organismo. A fin de concentrarnos en el asunto conceptual que nos interesa, supondremos que todo lo demás anda bien en el modo en que el mecanismo opera en estos casos.⁶ Nuestra segunda suposición concierne al concepto de “función”. Supondremos que el enfoque adecuado para pensar acerca de cómo se determinan las normas de función propia

⁶ Nótese que esto implica un grado relevante de simplificación. Por ejemplo, un asunto importante que no consideraremos aquí es que las personas con adicción a veces experimentan *anhedonia*, una falta de disfrute y de interés en la obtención de recompensas primarias. Es posible que esto sea causado por alteraciones en los parámetros normales del funcionamiento del sistema de recompensa (Volkow *et al.*, 2016, p. 366). Para una revisión general de la evidencia en torno a diferencias en estructura y función cerebral entre personas con y sin adicciones, cfr. Heilig *et al.* (2021). Nótese, adicionalmente, que no suponemos aquí que un abordaje de nivel neurobiológico sea necesariamente el adecuado para explicar la adicción. Para críticas a la suficiencia de las teorías neurobiológicas de la adicción, cfr. Flanagan (2025).

es, en un sentido amplio, evolucionista. Nos ubicamos, por decirlo rápidamente, entre los seguidores de Millikan (junto a Dennett), antes que entre los seguidores de Cummins. Lo que hace más interesante la discusión es, precisamente, que esto no alcanza para solucionar el problema al que hacemos referencia.⁷

Consideremos los dos puntos de vista contrapuestos. Primero: hay motivos para pensar que el sistema funciona correctamente, aun cuando el resultado sea la producción de deseos adictivos y estos sean tremendamente dañinos para los intereses y los planes de vida del agente. Tal es el modo usual de interpretar el punto central articulado por las teorías del desajuste: los resultados subóptimos no se deben a una falla propia del sistema de recompensa, sino al hecho de que este está operando bajo condiciones en cierto modo anómalas (*e. g.*, Durrant *et al.*, 2009; Ross, 2020). La intuición detrás de esta posición es que, en cierto sentido, el sistema está haciendo lo que se supone que debe hacer: reconocer la obtención de algo altamente gratificante y motivar al organismo a buscar y obtener más cosas de ese tipo.

Es importante en este contexto la distinción entre las condiciones ambientales actuales y lo que Tooby y Cosmides (1990, p. 386) llaman el “ambiente de adaptación evolutiva”, esto es, las condiciones ambientales en las que el sistema operaba cuando fue seleccionado. El sistema de recompensa, como cualquier otro mecanismo o rasgo biológico, fue seleccionado porque condujo confiablemente a resultados beneficiosos para la aptitud del organismo *bajo las condiciones en las que entonces operaba*. Crucialmente, el ambiente ancestral en el cual evolucionó el sistema de

⁷ Para resumir, Millikan (1984) sostiene que la función propia de un sistema debe identificarse apelando a actividades del sistema que permitieron adaptaciones relevantes para la supervivencia del organismo. David Papineau define esta posición como aquella según la cual “un rasgo T tiene la función F si fue diseñado por selección natural para producir F, por así decirlo – o, menos metafóricamente, si T está ahora presente es porque versiones ancestrales de T fueron seleccionadas porque produjeron el efecto F” (Papineau, 2017, p. 111; nuestra traducción). Cummins (1975), por su parte, sostiene que la función (propia) de un sistema debe identificarse por su contribución a sistemas de más alto nivel, sin importar ningún hecho pasado del sistema. Dennett se sitúa en la primera tradición, y con él asumiremos entonces una teoría teleológica de la función para propósitos de nuestro argumento. Sin embargo, uno de nosotros ha argumentado en otros trabajos en favor de teorías de la función al estilo de Cummins sobre otros fenómenos (Loaiza, 2024).

recompensa no incluía la presencia de drogas en cantidades suficientes para que los resultados negativos asociados a la adicción estuvieran entre las presiones selectivas bajo las cuales el sistema se formó.⁸ El resultado es un sistema que es, en efecto, vulnerable al “secuestro” por parte de sustancias que alteran directamente la disponibilidad de dopamina. Desde este punto de vista, la adicción es un resultado posible del hecho que el mecanismo no está diseñado para operar bajo tales condiciones —la evolución no contó con la oportunidad de diseñar una solución para ese problema—. ⁹ El punto importante, en relación con nuestra pregunta, es que no parece necesario suponer que el mecanismo como tal esté funcionando incorrectamente.

El punto de vista contrario articula también una idea importante y altamente intuitiva, a saber, que el mecanismo está fallando en su *razón de ser*, que no es otra que llevar al organismo a interesarse por las cosas que contribuyen a su aptitud biológica. Jerome Wakefield (2017a, 2017b y 2020) ha argumentado con fuerza a favor de esta posición. La intuición

⁸ Presumiblemente, no es hasta hace pocos miles de años que la disponibilidad de drogas en cantidades suficientes se volvió una posibilidad real, con la domesticación de las plantas y la aparición de la tecnología relacionada con la agricultura (Durrant *et al.*, 2009). Otra observación relacionada es que la adicción no parece ser un fenómeno que ocurra naturalmente en animales no humanos (fuera de condiciones de laboratorio, donde puede verse de forma muy clara), a pesar de que muchos de ellos cuentan con un sistema de recompensa similar al humano.

⁹ Considérese un razonamiento análogo para el caso de la obesidad, otro importante problema de salud en las sociedades contemporáneas que mayormente no existía en tiempos prehistóricos. La obesidad tiene consecuencias perjudiciales para la salud. Sin embargo, el mecanismo que nos motiva a ingerir tantas calorías como podemos es posiblemente un mecanismo que está funcionando bien, haciendo lo que está diseñado para hacer. El problema resulta de que el mecanismo fue diseñado para operar en un ambiente donde los alimentos ricos en calorías eran relativamente escasos, y donde, por tanto, podía ser adaptativa la estrategia de comer tanto como fuese posible cuando una fuente abundante se encontraba disponible. Sin embargo, esa estrategia no conduce ya confiablemente a resultados positivos en un entorno contemporáneo donde abundan los alimentos hipercalóricos. Es posible que muchos rasgos de los mecanismos que moldean nuestro comportamiento sean adaptaciones a circunstancias que ya no existen. La evolución biológica es simplemente demasiado lenta en comparación con el cambio explosivo en las condiciones de vida de los seres humanos en los últimos milenios.

central que los argumentos de Wakefield explotan es que, en casos de adicción, el sistema de recompensa está motivando al organismo a perseguir un (tipo de) objeto incorrecto, y no los que deberían ser, a la luz de su diseño, los *disparadores apropiados* de las señales de recompensa. Esta falla en cumplir su función propia etiológica es suficiente, según Wakefield, para que pensemos en la adicción como el resultado de un proceso disfuncional, aun cuando el mecanismo esté operando del modo previsto al formar deseos disposicionales a la luz de las señales de recompensa.

Creemos que ambas posiciones, la que atribuye el resultado subóptimo a las condiciones ambientales y la que atribuye una falla al mecanismo, capturan aspectos importantes del concepto de “función propia”. Al mismo tiempo, ambas motivan respuestas opuestas a la pregunta acerca de la presencia de una falla funcional en el sistema de recompensa en casos de adicción. A continuación, veremos cómo algunos aspectos del modo en que Daniel Dennett pensó acerca de las funciones biológicas contribuyen a la resolución de este problema. Presentaremos las ideas de Dennett en la sección siguiente, antes de volver a la pregunta en torno a la disfunción en la adicción en la sección 4.

3. Dennett sobre la subdeterminación y la función propia

La relevancia que posee el legado de Dennett en las investigaciones acerca de la intencionalidad —rasgo característico de algunos ítems (estados, eventos, sistemas, mecanismos, etc.) que son capaces de ser acerca de otros objetos, propiedades y relaciones en el mundo— es incuestionable. Nuestra invitación es a repensar este legado, al menos en lo que respecta a algunas cuestiones que son de importancia si deseamos evaluar críticamente la tensión que aparentemente existe entre *teorías de la disfunción* y *teorías del desajuste* a la hora de caracterizar qué sucede con el sistema de recompensa en el caso de la adicción a las drogas. Dentro de estas cuestiones hay una que nos parece que cumple un rol fundamental, a saber, el compromiso dennettiano con lo que entendemos es una tesis de subdeterminación de la teoría por parte de la evidencia empírica. Según esta tesis, existen múltiples descripciones o hipótesis distintas entre sí (por ejemplo, acerca de cuál es la función de un sistema) que son consistentes con la evidencia.

Para entender cómo es que Dennett llega a comprometerse con una tesis de subdeterminación de la teoría por parte de la evidencia empírica y cuál es su relevancia para nuestra discusión acerca de las adicciones es

importante que nos detengamos brevemente en tres ideas importantes para la filosofía dennettiana. La primera es de carácter metodológico/epistémico y apunta al hecho de que, al intentar comprender, explicar y predecir el comportamiento de cualquier sistema físico, habitualmente adoptamos una o más de las siguientes tres actitudes: la actitud intencional, la actitud de diseño y la actitud física (Dennett, 1978 y 1987). La actitud de diseño es la más relevante en el contexto de nuestra discusión acerca de cómo caracterizar la función propia del sistema de recompensa. Al adoptar la actitud de diseño *asumimos* que los sistemas han sido diseñados con un propósito y predecimos su comportamiento sobre la base de este sin necesidad de conocer los componentes físicos fundamentales del sistema. Esto nos permite ahorrar una gran cantidad de energía y tiempo, además de generar predicciones estadísticamente efectivas, siempre y cuando “el sistema al cual la estamos implementando se encuentre funcionando de manera propia (*proper function*) y se cumplan ciertas condiciones habilitantes” (Zawidzki, 2007, p. 36; nuestra traducción).

La importancia de la actitud de diseño nos lleva a considerar otra de las ideas dennettianas claves para nuestra discusión: la generalidad de las adscripciones funcionales. Para Dennett, las adscripciones funcionales que normalmente implementamos en el dominio de los artefactos no poseen características fundamentalmente diferentes de las que utilizamos en el caso de los seres humanos y otros sistemas biológicos (Dennett, 1987 y 1995). Tanto en el caso de los artefactos como en el de sistemas biológicos complejos, adoptamos una actitud en la que interpretamos las funciones propias de estos sistemas físicos *como si* estas respondieran necesariamente a las intenciones de un diseñador inteligente. Interpretamos, por ejemplo, la función propia de un radar o de un termómetro sobre la base de las supuestas intenciones de un agente inteligente que los ha diseñado, y algo análogo hacemos en el caso de los seres humanos, sus estados mentales y sus diferentes subsistemas biológicos. Al intentar comprender su función propia adoptamos una actitud en la que hacemos *como si* la selección natural — la “Madre Naturaleza” (Dennett, 1995, pp. 422-226)— tuviese intenciones capaces de satisfacer condiciones ideales de racionalidad óptima.

Adoptar la actitud de diseño y pretender que la función propia de los sistemas físicos responde a intenciones posee un valor instrumental heurístico. Sin embargo, más allá de este valor explicativo y predictivo, el legado del adaptacionismo darwinista que Dennett nos invita

a abrazar es otro: los organismos biológicos son el resultado de procesos algorítmicos, mecánicos, carentes de propósito, consciencia o inteligencia. La idea peligrosa de Darwin es un “ácido universal” (Dennett, 1996, p. 61) que nos permite comprender la emergencia de sistemas y fenómenos complejos sin explicaciones sobrenaturales o inteligentes. Los seres humanos, sus estados mentales y los subsistemas biológicos que los conforman caracterizados funcionalmente no son sino el resultado de procesos de selección natural que no requieren de un diseñador consciente con intenciones. Tal como afirma Dennett, “la ignorancia es una condición necesaria para muchas cosas excelentes” (1995, p. 514).

La generalidad de nuestras atribuciones funcionales y la ausencia de un diseño inteligente en los procesos de selección natural nos permite apreciar la importancia que posee para Dennett una última idea. Nos referimos a un supuesto de optimalidad, según el cual la comprensión de la función propia de cualquier sistema no solo requiere de la consideración de su historia evolutiva, sino también de lo que este sistema es capaz de hacer de la mejor forma dada su constitución física y su ambiente (Dennett, 1987 y 1995).¹⁰ Esto es sin duda un elemento importante de considerar a la hora de evaluar cuál es la función propia de cualquier sistema físico, incluyendo el sistema de recompensa en el contexto de la discusión acerca de la adicción a las drogas. Para ilustrar cómo es que los diferentes matices de la propuesta dennettiana contribuyen a establecer la subdeterminación empírica de la función propia de un sistema, revisemos un ejemplo paradigmático de su obra. Nos referimos al caso de la máquina expendedora de bebidas, equipada

¹⁰ La relación existente entre el supuesto de optimalidad y la adopción de una perspectiva histórica respecto de la función propia de un sistema nos permite comprender cómo es que un sistema ha maximizado su aptitud en un entorno dado. La selección natural tiende a privilegiar rasgos que contribuyen a mejorar las condiciones de reproducción y supervivencia de un organismo teniendo siempre en cuenta restricciones como la misma constitución física del sistema y el ambiente en el que se está desarrollando. La impronta dennettiana es que lo óptimo, es decir, lo que un sistema hace de la mejor forma en un momento determinado, siempre posee una explicación histórica. Para un interesante análisis de los modelos de optimalidad en biología y sus usos en las explicaciones por selección natural, cfr. Ginnobili y Roffé (2017).

con un dispositivo de dos bits para aceptar y rechazar monedas de veinticinco centavos de dólar (en adelante, *quarters-estadounidenses*):

Consideren una máquina expendedora de bebidas estándar, diseñada y construida en los Estados Unidos de América, y equipada con un dispositivo transductor para aceptar y rechazar monedas de veinticinco centavos de dólar (*quarters-estadounidenses*). Llamemos a este dispositivo un dispositivo de dos bits. Normalmente, cuando se inserta un *quarter-estadounidense* en un dispositivo de dos bits, el dispositivo entra en un estado, llamemos a este estado Q, que “significa” (observe las comillas irónicas) “yo percibo/acepto un *quarter-estadounidense* en este momento”. Esos dispositivos de dos bits son muy inteligentes y sofisticados, pero en ningún caso infalibles. Ellos “cometen errores” (más comillas irónicas). Esto quiere decir, de un modo no metafórico, que ocasionalmente ellos entran en el estado Q cuando se les introduce un trozo de metal u otro objeto extraño, y que a veces rechazan *quarters-estadounidenses* perfectamente legales: no entran en un estado Q cuando *se supone que deberían* hacerlo (Dennett, 1987, p. 290; nuestra traducción)

En este ejemplo, inspirado en un caso real, se considera un sistema físico que fue diseñado como un detector de *quarters-estadounidenses*. Al adoptar la actitud de diseño para comprender, explicar y predecir el comportamiento del dispositivo de dos bits, habitualmente apelamos a las intenciones de sus diseñadores, desarrolladores, propietarios y usuarios relativas a un ambiente específico. Le atribuimos una función propia al dispositivo en concordancia con estas variables indexadas a un ambiente y, sin embargo, sabemos que nuestras atribuciones son falibles. En ocasiones el dispositivo de dos bits entra en un estado de la clase Q en presencia de objetos extraños que, a pesar de no ser *quarters-estadounidenses* genuinos, son capaces de “engañar” al sistema. Si bien estos objetos son claramente diferentes a los *quarters-estadounidenses* en algunas de sus propiedades, evidentemente no lo son respecto de las propiedades que son relevantes a la hora de entrar en el estado Q como el peso, la forma, el diámetro o la composición material.

De hecho, en el caso específico del dispositivo de dos bits diseñado y construido en Estados Unidos que menciona Dennett, se constató que era posible utilizar monedas panameñas de veinticinco centavos para adquirir bebidas (en adelante: *quarters-panameños*). Esto llevó a que tiempo después estos dispositivos de dos bits fuesen instalados sin modificación alguna en máquinas expendedoras de bebidas en Panamá para ser utilizados como detectores de *quarters-panameños*. Por ello se hace referencia a esta clase de dispositivo de dos bits como “errante” (Dennett, 1987, p. 290). ¿Cómo determinar la función del dispositivo de dos bits? Sabemos que el dispositivo fue diseñado por seres humanos en un contexto o ambiente determinado. Esto permite, en un primer momento, determinar su función como la de un *detector-de-quarters-estadounidenses* y no disyuntivamente como la de un *detector-de-quarters-estadounidenses* o un *detector-de-quarters-panameños*. Aquí la “función propia” del dispositivo se entiende sobre la base de las intenciones de sus diseñadores en Estados Unidos y es solo en virtud de la función propia comprendida así que podemos referirnos al dispositivo como funcionando correcta o incorrectamente.

Ahora bien, señalamos que tiempo después el mismo dispositivo fue instalado sin alteración alguna en Panamá, de modo tal que sus usuarios pudiesen adquirir bebidas usando *quarters-panameños*. Teniendo en consideración este nuevo nicho o ambiente surgen nuevas preguntas (Dennett, 1987, pp. 292-293): ¿tiene el dispositivo de dos bits la misma función en este nuevo nicho? Si adoptamos la estrategia dennettiana nos daremos cuenta de que cualquier intento por determinar la función propia de este dispositivo (o de cualquier otro sistema) no solo debe integrar consideraciones históricas con relación a los procesos de selección de sus rasgos, sino también consideraciones pragmáticas asociadas a la utilidad que este sistema tiene en su nuevo nicho o ambiente (Dennett, 1987, p. 293). La razón de ser de un sistema, el dispositivo de dos bits en este caso, va a estar siempre indexada a la utilidad que ese sistema o dispositivo tiene (tuvo) en el nicho específico en el que opera (operaba), proporcionando así una ventaja adaptativa. En consecuencia, no hay un hecho decisivo acerca del sistema o dispositivo que nos permita determinar su función propia con independencia de su historia y de la utilidad que posee en el nicho o ambiente en el que de hecho se encuentra operando.

A nuestro parecer, sin embargo, el alcance de la heurística dennettiana es más radical. Uno podría interpretar a Dennett como

alguien que simplemente tiene como objetivo la defensa de una variante de externalismo acerca de la atribución funcional, en cuanto la especificación de la función propia de un sistema depende al menos en parte de hechos externos como el ambiente y la intención de los diseñadores y/o usuarios. No obstante, más allá de si los hechos en cuestión son internos o externos al sistema, es importante recordar que Dennett niega enfáticamente que exista algo como un hecho decisivo (*fact of the matter*) sobre el sistema cuyo comportamiento estamos intentando comprender, explicar y predecir que nos permita determinar unívocamente cuál es “realmente” su función propia o el contenido de sus estados con independencia de diversas consideraciones pragmáticas asociadas a la actitud que nosotros adoptamos al comprender, explicar y predecir.

La invitación de Dennett está amparada en la adopción de una tesis acerca de la subdeterminación de la evidencia empírica (Quine, 1951 y 1975) que parece derivarse de las consideraciones anteriormente expuestas. Las atribuciones funcionales *qua* hipótesis estarían subdeterminadas empíricamente en el sentido de que para el mismo conjunto de evidencia empírica pueden existir múltiples hipótesis capaces de explicar adecuadamente la evidencia. No hay un hecho decisivo que nos permita saldar la cuestión acerca de cuál de estas hipótesis es la verdadera a partir de elementos intrínsecos al sistema, esto es, más allá de las consideraciones acerca de su —y de criterios de— utilidad. Pensemos nuevamente en el dispositivo de dos bits. En este caso disponemos de diferentes descripciones que son consistentes con la evidencia empírica a pesar de ser distintas entre sí. Por ejemplo, cuando el dispositivo operaba inicialmente en Estados Unidos, ¿existía algún hecho decisivo intrínseco acerca del dispositivo que nos permitía atribuirle la función propia de un *detector-de-quarters-estadounidenses* y no otra diferente consistente con la evidencia? Una vez en Panamá, ¿qué nos permite asegurar que el dispositivo opera correctamente solo cuando funciona como un *detector-de-quarters-panameños*?

La subdeterminación empírica parece acarrear formas de relativismo ontológico donde nuestro entendimiento de lo que existe es relativo al marco teórico o actitud científica que adoptamos (De Haro, 2023). Interpretamos los sistemas *como si* tuviesen una función propia que no depende de nuestras prácticas epistémicas. ¿Son entonces estas funciones algo que solo proyectamos y no algo cuya realidad reside en el mundo natural? Aquí debemos ser extremadamente cautos. Por una parte,

Dennett claramente se desmarca de realismos “históricos” (2017, p. 14) que poseen “fuerza industrial” (1991, p. 32) en el caso de las atribuciones funcionales. Este realismo estaría comprometido con la existencia de propósitos, significados e intencionalidad en términos ontológicamente robustos, es decir, propósitos cuya determinación es independiente de cualquier clase de consideración indexada a nuestra perspectiva epistémica. Por otra parte, Dennett tampoco se siente cómodo con un mero instrumentalismo en el que toda función o propósito no es más que una metáfora con valor predictivo. Por ello él insiste en calificar su enfoque como una forma de “realismo moderado” (Dennett, 1991, p. 30) en el que la intencionalidad y los propósitos sí existen en la medida en que estos “pueden ser discernidos solo desde el punto de vista de alguien que adopta cierta estrategia predictiva” (Dennett, 1987, p. 15). ¿Apuntamos a algo real cuando realizamos atribuciones funcionales? ¿Poseen *realmente* el dispositivo de dos bits o el sistema de recompensas una función propia? El realismo moderado dennettiano es escueto, pero sostiene que cualquier persona que intentara comprender el comportamiento humano o cualquiera de sus subsistemas sin atribuir intenciones o funciones estaría perdiéndose de un patrón real en los datos.

En síntesis, la propuesta dennettiana conlleva un compromiso con una tesis de subdeterminación de la teoría por parte de la evidencia empírica, en este caso sobre la función propia de los sistemas biológicos. Este compromiso implica que no hay un hecho decisivo intrínseco que nos permita dirimir cuál de las hipótesis o descripciones empíricamente adecuadas en juego son correctas. A continuación, evaluaremos las consecuencias de estas consideraciones para el caso específico de la disputa acerca de la función propia del sistema de recompensa.

4. Subdeterminación y la función propia del sistema de recompensa

Podemos ahora enunciar sucintamente nuestra tesis central: la situación del sistema de recompensa en la adicción es análoga a la del dispositivo de dos bits luego de su traslado a Panamá. La máquina expendedora fue equipada con un mecanismo que le permite reconocer objetos de un tamaño y un peso específicos, en un contexto donde ese tamaño y peso específicos covarían confiablemente con la propiedad de ser un *quarter-estadounidense*. Luego de la aventura histórica que llevó

a la máquina a las costas panameñas, ese mismo mecanismo pierde la capacidad de indicar de modo confiable la propiedad de ser un *quarter-estadounidense*, ya que los *quarters-panameños* tienen un tamaño y un peso que activan el mecanismo exactamente del mismo modo en que los *quarters-estadounidenses* lo hacen. Por un lado, hay un sentido en el que la máquina está funcionando bien, su sistema de detección de objetos no se ha visto alterado y no se encuentra descompuesto. Al mismo tiempo, parece razonable suponer que es la capacidad del mecanismo de indicar confiablemente la presencia de monedas estadounidenses, y no la de monedas panameñas, la que explica por qué el mecanismo existe, la que determina su razón de ser. En consecuencia, hay otro sentido en el que parece razonable decir que la máquina está fallando, dado que no está cumpliendo con su cometido.

Presumiblemente, la razón de ser del sistema de recompensa es su capacidad de generar deseos disposicionales intrínsecos por cosas valiosas para el organismo desde el punto de vista biológico. Cuando la operación del sistema resulta en la formación de deseos disposicionales intrínsecos por consumir drogas, resultando una fuente de motivación extraordinariamente fuerte y persistente, hay un sentido en que es razonable decir que el sistema está fallando —no está cumpliendo con el propósito para el cual fue diseñado; de hecho, puede que esté obstaculizando la consecución de ese propósito—. Al mismo tiempo, hay un sentido en que el sistema continúa detectando correctamente el parámetro que, se supone, se espera que sea capaz de detectar: la correlación entre ciertos eventos y episodios de aumento en la disponibilidad de dopamina. El sistema mantiene su capacidad de detectar tales correlaciones y aprender de ellas del modo esperado. De acuerdo con esta descripción, el sistema está funcionando correctamente mientras que, de acuerdo con la otra, el sistema está fallando.

¿Cuál de las dos descripciones es correcta? Parece imposible decidir si un sistema está funcionando incorrectamente sin tener una idea clara respecto de qué es lo que el sistema *debería* estar haciendo, si estuviese operando correctamente. Es precisamente en este punto donde se ubica el desacuerdo fundamental entre las teorías de la disfunción y las teorías del desajuste. De acuerdo con las primeras, el sistema se aparta de su funcionamiento normal en la medida en que propicia que el organismo persiga los objetos inadecuados. De acuerdo con las segundas, el sistema puede ser visto como funcionando normalmente en la medida en que tenga éxito en la identificación de objetos que covarían con aumentos

en la disponibilidad de dopamina. Así, la discusión entre ambas teorías es en buena medida una discusión acerca de cuál de estas dos hipótesis nos da una descripción adecuada de la función propia del sistema de recompensa.

Es exactamente en este punto que las ideas de Dennett juegan un papel importante. Consideremos el dispositivo de dos bits en primer lugar. ¿Es su función detectar *quarters-estadounidenses* o detectar objetos con ciertas características específicas de peso y tamaño? El dispositivo de dos bits es un artefacto en el sentido corriente del término, diseñado por una o más personas que perseguían un propósito al diseñar y construir una máquina con esas características. Dada una concepción evolucionista de la función, la función del sistema está determinada por ciertos hechos históricos, específicamente hechos relativos al propósito con el que fue diseñado. Dado que sus diseñadores son personas en el sentido corriente del término, es posible, al menos en principio, preguntarles acerca del fin que perseguían al diseñar y construir una máquina de esas características; si esto no fuese posible, podemos consultar los documentos con las especificaciones de diseño de la máquina, o intentar obtener por otros medios evidencia acerca de las intenciones de sus diseñadores. El proceso dista de ser infalible, como bien sabe la arqueología. Sin embargo, para el caso de la máquina expendedora, podremos seguramente formular hipótesis plausibles y llegar a una respuesta suficientemente convincente respecto de cuál es el propósito que sus diseñadores perseguían.¹¹

Consideremos ahora el sistema de recompensa. ¿Es su función formar deseos por cosas biológicamente valiosas, o formar deseos por cosas que covaríen confiablemente con episodios de aumento en la disponibilidad de dopamina? Adoptar la actitud de diseño nos permite entender muchas cosas respecto del mecanismo y su función. Si nos preguntamos para qué es el sistema de recompensa, la respuesta hemos de buscarla en hechos históricos acerca de cómo es que llegamos a tener

¹¹ Como señalamos en la sección anterior, el asunto es, en realidad, ligeramente más complicado, ya que las intenciones de otros agentes más allá de los diseñadores juegan también un papel en la explicación de cómo es que la máquina llegó a estar equipada con un dispositivo de ese tipo. En particular, juegan un papel allí las intenciones de quienes decidieron trasladar la máquina a Panamá, quienes determinaron subsiguientemente que continuara en operación, etc.

un mecanismo de este tipo que nos revelan algo acerca de su razón de ser. A ciertos fines, es útil atribuir intenciones a su “diseñadora”, la Madre Naturaleza, y preguntarnos qué es lo que ella pretendía cuando ideó un mecanismo con ciertas características determinadas —o cuando cooptó un rasgo o mecanismo previamente existente para un fin diferente al original—. Sin embargo, no podemos preguntarle a la Madre Naturaleza cuáles eran exactamente sus intenciones, ni contamos con documentos que detallen especificaciones de diseño con base en los cuales determinar si la función del mecanismo era formar deseos por cosas biológicamente valiosas o por cosas que incrementen la dopamina. Lo que sabemos es, a grandes rasgos, que hubo en algún momento distante en el tiempo animales que llegaron a tener un sistema con ciertas características determinadas y que tener un sistema de este tipo provocó una serie de efectos que dieron como resultado una mejora apreciable en la aptitud biológica de sus portadores. La observación crucial es que ambas hipótesis acerca de la función del sistema de recompensa —como un sistema para desear cosas biológicamente valiosas, o como un sistema sensible al aumento en la disponibilidad de dopamina— son *igualmente* compatibles con tales hechos acerca de cómo llegó a formarse el sistema, y esos hechos son todo lo que tenemos para constreñir el espacio de hipótesis adecuadas acerca del fin perseguido por la Madre Naturaleza. Las explicaciones que ofrecen ambas hipótesis acerca del modo en que el sistema contribuye a aumentar la aptitud biológica son empíricamente equivalentes. En otros términos, la evidencia deja subdeterminada la elección entre ambas hipótesis.

Ahora bien, la existencia de cierto grado de subdeterminación no implica que cualquier hipótesis sea correcta o que no haya ninguna hipótesis científica útil por formular. Las consideraciones históricas, combinadas con la actitud de diseño, nos permiten realizar conjeturas acerca de las “intenciones” de la Madre Naturaleza que pueden iluminar aspectos importantes del proceso que condujo a la formación del mecanismo. Podemos, sobre esta base, formular explicaciones y predicciones fructíferas en muchos sentidos y descubrimos, en el proceso, patrones que de otro modo serían difíciles de advertir. En este sentido, cuando nos preguntamos por la función de un rasgo o mecanismo, hay hechos que nos permiten constreñir significativamente el espacio de respuestas posibles; no todas las hipótesis son compatibles con los hechos. Sin embargo, sería un paso en falso suponer que, en el caso que nos ocupa, *ha de haber* algún hecho decisivo que nos permita

resolver la pregunta acerca de cuál de las hipótesis alternativas acerca de la función del sistema de recompensa es correcta. La evidencia disponible es igualmente compatible con ambas hipótesis. La atribución de intenciones a la Madre Naturaleza es un expediente útil a ciertos fines, pero no es una herramienta que nos permita definir fructíferamente una cuestión como esta.

El acercamiento que proponemos al debate entre teorías de la disfunción y teorías del desajuste, al menos en cuanto a las adicciones, es *disolutorio*. Los hechos relevantes para explicar por qué existen organismos dotados de un sistema con estas características son igualmente compatibles con ambas descripciones de la función del sistema de recompensa. Y no hay ningún hecho ulterior al que podamos recurrir para definir entre las dos hipótesis en pugna. Hemos propuesto que el *quid* de la discusión entre ambos enfoques reside en que estos operan con hipótesis diferentes acerca de la función propia del sistema de recompensa, y hemos argumentado que la elección entre ambas hipótesis está subdeterminada por la evidencia. Advertir esto es un antídoto necesario contra la intuición de que ha de haber algún *hecho decisivo* que nos permita elegir entre ambos tipos de teorías. Si nuestro argumento es correcto, no hay tales hechos decisivos. Como veremos en lo que sigue, sin embargo, esto no implica que la inclinación por una teoría de uno u otro tipo sea necesariamente arbitraria.

5. Objeciones e implicaciones

En esta sección, consideraremos algunas objeciones al enfoque dennettiano que resultan relevantes para nuestro argumento. Específicamente, nos concentraremos en críticas a la manera en la que Dennett piensa en la identificación de la función propia de un sistema y su compromiso con la idea de que la descripción de la función propia está subdeterminada por la evidencia. Nos concentraremos en la objeción según la cual aceptar algún grado de subdeterminación de la función propia del sistema de recompensa no es estrictamente una solución, sino rendirse ante el problema. Según esta objeción, aceptar que no hay hechos que determinen unívocamente la función propia de un sistema (o que permitan escoger una única hipótesis como la correcta) equivale a negarse prematuramente a identificar hechos del mundo que sí resultan relevantes para la identificación de funciones.

Una versión de esta objeción puede formularse atendiendo a las observaciones de Garson (2019) sobre la subdeterminación de las

funciones. Garson sostiene que aceptar sin más la subdeterminación de la función es problemático por tres razones. La primera es que nos impide hablar de disfunciones, pues cualquier atribución de disfunción exige especificar el funcionamiento correcto de un sistema, lo que, dado un enfoque etiológico sobre la función, supone estudiar qué efectos de hecho han aumentado la aptitud biológica del organismo. La segunda es que llevaría a la indeterminación del contenido, pues sería imposible hablar de algún contenido adecuado o inadecuado en el caso de sistemas cuya función es representar.¹² Finalmente, no determinar correctamente la función propia de un mecanismo podría llevar a intervenciones inadecuadas e incluso peligrosas. El ejemplo que Garson considera es la hipótesis según la cual la esquizofrenia se debe a un exceso de dopamina en el cerebro, lo que invitó a intervenir únicamente sobre los niveles de dopamina de pacientes con esquizofrenia, ignorando otros factores que resultan relevantes para su tratamiento adecuado y provocando efectos secundarios indeseables.

Estas objeciones pueden aplicarse de manera clara al argumento que ofrecemos sobre la adicción. En primer lugar, tendríamos serias dificultades para distinguir disfunción y función propia del sistema de recompensa en el contexto de la adicción. En el peor de los casos, no podríamos hacer distinción alguna. En el mejor de los casos, podríamos hacer algunas distinciones (dentro de un campo restringido de hipótesis), pero estas no serían suficientes para los propósitos explicativos ni prácticos de la psiquiatría. Segundo, al no poder determinar funciones ni disfunciones, tampoco podemos establecer qué contenido representacional sería adecuado o inadecuado para el sistema de recompensa. Suponiendo que las representaciones de este sistema evalúan un objeto como deseable, bueno o valioso, no podríamos decir que sea inadecuado para el sistema evaluar así ciertos objetos (entre ellos las drogas). Finalmente, esto puede impedirnos formas de intervención, pues podríamos pensar equivocadamente que podemos intervenir en puntos del sistema en donde la intervención es fútil (*e. g.*, intervenir sobre la visión para evitar que una persona vea sustancias en su ambiente).

Para responder a esta objeción, comenzaremos por notar lo que consideramos el espíritu central de esta. La motivación general de la

¹² Estos dos puntos pueden encontrarse también en Fodor (2000).

objeción, a nuestro modo de ver, es un compromiso con el realismo y la expectativa de que, acumulando más evidencia, podremos resolver la discusión. Esta motivación es razonable, pues un enfoque naturalista como el que estamos adoptando debe evaluarse sobre la evidencia disponible en cada momento. No obstante, creemos que podemos conservar nuestro enfoque naturalista y la sensibilidad a la evidencia, manteniendo el argumento que hemos defendido sobre la identificación de la función propia del sistema de recompensa. A continuación, mostraremos cómo aceptar que la elección entre hipótesis sobre la función propia del sistema está empíricamente subdeterminada es compatible con enfoques que sean sensibles a la acumulación y descubrimiento de nueva evidencia.

En primer lugar, el argumento de la subdeterminación no implica que *ningún* hecho del mundo sea relevante para la identificación de funciones en general. En los casos que presenta Dennett, por ejemplo, hay hechos del mundo que restringen el conjunto de descripciones aceptables de la función del dispositivo de dos bits. La pregunta sobre este mecanismo es si el dispositivo tiene como función propia detectar monedas estadounidenses o panameñas, no si su función es detectar metales o sólidos. Aunque la descripción de su función está subdeterminada en algún grado, hay funciones que claramente están excluidas del rango de hipótesis a considerar. Estas restricciones se imponen por hechos sobre la constitución física del mecanismo (*e. g.*, el dispositivo es incapaz de detectar gases, por lo que su función no puede ser la detección de ningún gas) o por su historia de diseño (*e. g.*, el dispositivo fue llevado a Panamá, por lo que su función no puede ser detectar pesos chilenos), de modo que hay al menos algunos hechos que resultan relevantes para restringir las posibles funciones que un mecanismo o sistema puede tener.

Lo que creemos que podemos aprender de Dennett, específicamente, es que no es necesaria una especificación de las funciones de los sistemas (o del contenido de la representación) *más allá* de cierto punto en el que la explicación comienza a predecir y generalizar de manera productiva. En otras palabras, no es que no podamos hablar de funciones en absoluto, sino que no debemos pensar que solo habrá una única hipótesis posible que sea empíricamente adecuada y a la que llamemos la descripción verdadera de la función propia de un sistema. Es en este sentido que no hay hechos en el mundo que permitan determinar la función, esto es, que permitan dar con una única descripción verdadera de la función del

sistema. En términos más cercanos a Quine (1951), el punto central es que dentro de un rango de hipótesis que resultan empíricamente adecuadas, determinar cuál hipótesis adoptar ya no responderá a criterios empíricos o lógicos, sino a condiciones pragmáticas de la explicación —un asunto sobre el que volveremos en un momento—. La elección de hipótesis está subdeterminada por la evidencia, pero esto dentro de un conjunto de hipótesis que ya hemos identificado como empíricamente adecuadas, no frente a cualquier hipótesis lógicamente posible.

Dada esta aclaración de nuestro argumento, podemos responder a la objeción de dos maneras. Primero, la objeción exige algo que es imposible o al menos innecesario al pedirnos no renunciar a la búsqueda de una única descripción de la función propia de un sistema. Por un lado, si aceptamos los argumentos de Quine, no hay consideración lógica que permita escoger una única hipótesis como empíricamente adecuada, incluyendo hipótesis sobre la función propia. Por otro lado, incluso si rechazamos el argumento quineano, encontrar una única descripción de la función sería pragmáticamente innecesario, toda vez que es posible encontrar descripciones que sean productivas aun aceptando algún grado de subdeterminación empírica. En este sentido, la ciencia puede avanzar dentro de un espectro de descripciones adecuadas. Basta con tener claridad sobre los objetivos epistémicos y prácticos de la explicación para determinar el grado de subdeterminación aceptable, pero no será necesario eliminar todo grado posible de subdeterminación.

Segundo, sostener alguna forma de realismo, como pretende la objeción, no requiere necesariamente comprometerse con la adopción de una descripción unívoca de la función propia. El argumento de la subdeterminación que estamos defendiendo, como lo entendemos, sostiene que no es necesaria una única hipótesis de la función de un sistema para generar hipótesis productivas, pero, como hemos argumentado, esto es compatible con que haya hechos relevantes en la delimitación de las hipótesis. Esto se muestra en el caso del dispositivo de dos bits, donde hay hechos que delimitan el espacio de hipótesis aceptables acerca de su función. Es posible también que avances futuros cambien nuestros requisitos de aceptabilidad de alguna descripción de estas funciones. Por ejemplo, dada nuestra tecnología actual y el uso cada vez más infrecuente de monedas, lo que antes era un detector de monedas podría acabar siendo un detector de coleccionables. En este sentido, creemos que hay algo que podemos llamar “real” al apostarle a alguna hipótesis sobre la función de un sistema, pues hay hechos del

mundo que hacen que algunas hipótesis sean mejores descripciones del mundo que otras. No obstante, esto no implica que debamos encontrar una sola hipótesis correcta, pues podemos tener distintas maneras de describir y conocer la realidad. Para volver a Dennett, podemos ver patrones reales desde ciertas perspectivas y no desde otras, sin que esto signifique que alguno de estos patrones no sea real.¹³

No obstante nuestra defensa, hay también lecciones que extraer de las objeciones que hemos considerado. Si bien podemos aceptar cierto grado de indeterminación sobre el contenido y la función propia a la hora de discutir sobre los mecanismos involucrados en la adicción, quedan sin especificar los criterios de éxito (pragmáticos o no) de una explicación adecuada de este fenómeno. En otras palabras, incluso si aceptamos la subdeterminación dentro de un rango restringido de hipótesis empíricamente equivalentes, ¿cuáles son entonces los criterios de adopción de una u otra hipótesis? Responder a esta pregunta yace por fuera del espacio de este artículo, pero podemos entrever algunas ideas interesantes que pueden motivar discusiones productivas sobre la adicción.

Una reacción interesante a las tesis quineanas sobre la subdeterminación de las teorías por la evidencia es pensar que la elección entre hipótesis empíricamente equivalentes puede hacerse sobre la base de criterios de valor. En el trabajo de Quine, los criterios que aparecían invocados para seleccionar entre estas hipótesis eran criterios “pragmáticos”, pero no es muy claro qué cuenta como un criterio pragmático de elección. Intuitivamente, podemos invocar aquí valores epistémicos como la simplicidad o la consistencia, es decir, preferimos

¹³ Según Dennett (1991), distintas perspectivas permiten ver patrones a distintos niveles, todos los cuales tienen derecho legítimo a ser llamados reales. Adoptar distintas perspectivas, actitudes o *stances*, en el marco más amplio de la filosofía de Dennett, permite notar distintas formas en las que el mundo es, sin abandonar la idea de que estamos describiendo algo que llamamos real. Aplicado al caso de las funciones, el punto central es que, aunque tengamos varias descripciones de cuál es la función de un sistema (e. g., si el sistema de recompensa tiene como función motivar la producción de dopamina o la búsqueda de recompensas naturales), es posible hablar de ellas como descripciones de una función real, *i. e.*, de un patrón que describe la realidad del sistema. En consecuencia, el enfoque dennettiano no invita a abandonar los compromisos realistas que la objeción busca rescatar, y creemos que es posible hacer compatibles ambas ideas de forma productiva.

hipótesis que maximicen la consistencia con el resto de nuestro sistema de creencias y que postulen la menor cantidad posible de entidades. Estos valores se basan en la idea de que el objetivo de adoptar alguna hipótesis es explicar y predecir fenómenos en el mundo de manera compatible y bien integrada con el resto de nuestro conocimiento.

A pesar del protagonismo que han tenido criterios de elección como los anteriores, hay otros valores que pueden contribuir a la elección de descripciones e hipótesis en psiquiatría. Filósofas feministas de la ciencia como Harding (1986) y Longino (1990) han mostrado cómo los criterios pragmáticos de elección entre hipótesis involucran consideraciones de valor que pueden incluir valores no epistémicos. Entre estos valores, recalcan la presencia de valores dependientes de estructuras de género, aunque pueden existir muchos otros tipos de valor. Entre los casos que estudian se encuentran la consideración de hipótesis en biología, arqueología, e incluso física y matemáticas. Según estas filósofas, qué hipótesis son consideradas, priorizadas y justificadas, o modificadas y potencialmente rechazadas, depende no solo de consideraciones lógicas y empíricas, sino de los compromisos éticos e incluso políticos que los investigadores y las investigadoras mantengan. Esto ocurre particularmente en casos de decisiones bajo subdeterminación (Carrier, 2011).

La idea de que consideraciones acerca de valores pueden jugar un papel relevante ante un escenario de subdeterminación es consistente con una visión más amplia de la psiquiatría como una disciplina donde los criterios valorativos juegan un papel destacado. Tal visión no es de ningún modo nueva (Kostko, 2019). Quienes defienden la presencia de valores en la psiquiatría han subrayado, por ejemplo, que la APA evalúa modificaciones al DSM en parte sobre consideraciones relativas al daño potencial de la creación, modificación o eliminación de categorías (Cooper, 2004)¹⁴. Esto implicaría que la clasificación estándar en psiquiatría, y en consecuencia la identificación de trastornos psiquiátricos, no involucra exclusivamente consideraciones epistémicas. Adicionalmente, hay buenas razones para pensar que la presencia de valores en psiquiatría ha llevado a encontrar evidencia relevante para el estudio de fenómenos como el trastorno del espectro autista y abandonar líneas de investigación estancadas (Ahn, 2020). Para la

¹⁴ Documentos recientes sobre estos procesos en la modificación del DSM pueden encontrarse en APA DSM Team (2021).

discusión filosófica, la conclusión que extraen estos filósofos y filósofas es que una descripción adecuada de la ciencia debe considerar los usos de hecho de los valores en la práctica científica.

Los argumentos usuales a favor de la presencia de valores en ciencia cobran un cariz especial en la psiquiatría, al tratarse de una disciplina cuyos objetivos no son únicamente científicos, sino también médicos. En otras palabras, el objetivo de la psiquiatría no es únicamente comprender mejor los fenómenos mentales, sino intervenir y tratar posibles afectaciones. Esto implica que la adopción de una hipótesis en psiquiatría no responde únicamente a valores epistémicos, sino también a consideraciones prácticas, esto es, a los cursos de acción que la hipótesis habilita. Finalmente, estos cursos de acción serán evaluados en parte por su efectividad, pero también por su concordancia con valores éticos de la práctica médica. En consecuencia, la elección de hipótesis en psiquiatría parece sensible a cuestiones de valor en varios sentidos, implicando valores epistémicos, clínicos y ético-políticos. De aceptar esto, entonces creemos que nuestro argumento tiene consecuencias muy fructíferas para explorar.¹⁵

Consideremos, por ejemplo, los debates sobre la clasificación de la adicción como trastorno (Burdman, 2025). Intuitivamente, sostener que la adicción involucra disfunción en el sistema de recompensa promovería la identificación de la adicción como trastorno. Sin embargo, por contrapartida, inclinarse a no clasificar la adicción como trastorno invitaría a no caracterizar el sistema de recompensa como disfuncional en este contexto. Si aceptamos ahora que no hay un hecho decisivo que nos permita establecer que el sistema de recompensa es disfuncional en la adicción, se sigue entonces que la discusión sobre la disfunción no tiene prioridad sobre la discusión de la clasificación de la adicción como

¹⁵ ¿Podría suceder que la elección entre teorías quedara subdeterminada incluso luego de añadir la consideración de factores pragmáticos y valorativos? Aunque parece difícil imaginar ese escenario como un desafío práctico concreto para la elección entre teorías, admitimos que seguramente no sea inconcebible. Nótese, sin embargo, que esto no supone que la evidencia empírica y las consideraciones valorativas o pragmáticas estén a la par como criterios de elección entre teorías, ya que las segundas solo entran en juego, de acuerdo con nuestra propuesta, como criterios de elección entre teorías empíricamente equivalentes. Agradecemos a un referí anónimo por alentarnos a clarificar este punto.

trastorno y los valores que juegan un papel allí. Ambas discusiones, entonces, van de la mano, una consecuencia interesante para quienes defienden el rol de los valores en la psiquiatría.

Además de los asuntos sobre la clasificación, una consecuencia interesante de este argumento es que nos permite enfocar mejor la discusión en torno a la adicción y sus mecanismos subyacentes. Hasta el momento, la discusión entre teorías del desajuste y teorías de la disfunción se ha concentrado en la pregunta por la disfunción. Si nos desprendemos de esta última pregunta aceptando el argumento disolutorio, podemos redireccionar la discusión a, por ejemplo, comprender y proponer distintos modos de intervención. En el caso del sistema de recompensa, a modo de ejemplo, ver el sistema bajo la hipótesis de que su función es la búsqueda de recompensas naturales (y en consecuencia califica como disfuncional en el caso de la adicción, como sostienen las teorías de la disfunción) nos permite identificar potenciales vías de intervención a nivel cerebral orientadas a intervenir sobre los circuitos de dopamina. Esto promovería intervenciones como el uso de estimulación magnética transcraneana para modular la producción de neurotransmisores involucrados en la adicción (Gorelick *et al.*, 2014). Sin embargo, ver ahora el sistema bajo la hipótesis de que su función es responder a aumentos en la disponibilidad de dopamina (y en consecuencia no califica como disfuncional, siguiendo a las teorías del desajuste) nos permite entender por qué también es importante intervenir sobre el ambiente para evitar el “secuestro” del sistema por el consumo de sustancias. Esto puede resultar deseable, especialmente al ofrecer vías de intervención no invasiva y más efectivas a largo plazo. En este orden de ideas, distintas descripciones del sistema promueven distintas intervenciones y modos de acción. El resultado de nuestro argumento es entonces no solo epistémicamente fructífero, al enfocar mejor los esfuerzos científicos en el diseño de distintas intervenciones, sino éticamente deseable, al traer al centro del escenario discusiones urgentes sin tropezar con los problemas conceptuales en los que el debate se ha concentrado.

Ahora bien, si bien hemos tomado el ejemplo de la discusión sobre distintos modos de intervención, nuestro argumento es más general. A nuestro modo de ver, dejar atrás la discusión sobre la presunta disfunción del sistema de recompensa nos permite desplazarla a la pregunta sobre cuáles son nuestros intereses cuando estudiamos este sistema y su relación con la adicción. Estos intereses cobijan el interés de intervenir

a nivel individual, pero también otros posibles intereses sobre cómo y dónde intervenir para enfrentar toda clase de problemas asociados, tales como problemas de salud y política pública, aspectos sociales de la adicción, estrategias preventivas, entre otros. Si bien decidir qué intereses deben jugar un papel importante en la consideración de hipótesis sobre el sistema de recompensa yace por fuera de este artículo, creemos que nuestro argumento da pasos importantes para llevar la discusión a este plano. Esto resulta razonable desde nuestra perspectiva, y especialmente llamativo para quienes ya aceptan que las consideraciones de valor deben ser parte crucial de los debates en psiquiatría.

6. Conclusión

En un primer acercamiento, el contraste entre teorías de la disfunción y teorías del desajuste parece reflejar un desacuerdo fundamental acerca de los hechos relevantes para comprender los trastornos mentales. Según hemos argumentado en las secciones anteriores, esta impresión es equivocada, al menos en la versión particular de este debate que concierne a la adicción. Hemos propuesto que el centro del desacuerdo entre ambos tipos de teorías se ubica en dos visiones contrapuestas de la función propia del sistema de recompensa. Las teorías de la disfunción asocian la función de este sistema al reconocimiento de ítems biológicamente valiosos y la consecuente motivación por obtenerlos, mientras que las teorías del desajuste caracterizan su función en términos de la detección de ítems que provocan aumentos en la disponibilidad de dopamina y la consecuente motivación por obtenerlos. Hemos propuesto que adoptar el modo dennettiano de pensar acerca de la función propia resulta un poderoso remedio frente a la impresión de que el asunto en discusión es empíricamente significativo. La atribución de funciones biológicas supone, desde la perspectiva dennettiana, la atribución de intenciones a la Madre Naturaleza, la cual se ve constreñida por hechos históricos —concretamente, hechos acerca de los efectos que se siguieron de la posesión de los rasgos relevantes en el ambiente de adaptación evolutiva—. Sin embargo, ante dos hipótesis que son igualmente compatibles con tales hechos, la elección entre ambas se encuentra subdeterminada y no hay ningún hecho adicional que funcione como hecho decisivo para dirimir la cuestión. La atribución de funciones biológicas, como la atribución de intenciones a la Madre Naturaleza, es un expediente útil a ciertos fines, pero no uno que nos permita resolver una cuestión como esta. Recurrir a esta visión de la función propia nos

permite entonces disipar la impresión de que el contraste entre teorías de la disfunción y teorías del desajuste refleja una diferencia significativa acerca de los hechos relevantes, lo que nos ha llevado a proponer una salida disolutoria frente a este debate. Creemos, al mismo tiempo, que se sigue de ello una ganancia positiva. Dejar atrás la impresión de que la elección entre ambos tipos de teorías depende de asuntos empíricamente discernibles contribuye a redirigir nuestra atención a otro tipo de consideraciones que se ven oscurecidas por el modo en el que el debate entre ambas es usualmente enfocado. En particular, hemos sostenido que la relectura propuesta de esta discusión nos permite ver con más claridad el papel que en ella juegan diferentes consideraciones pragmáticas y valorativas. Esto, según creemos, conduce el debate hacia un terreno más fructífero.¹⁶

Referencias

- Ahn, S. (2020). How Non-Epistemic Values Can Be Epistemically Beneficial in Scientific Classification. *Studies in History and Philosophy of Science Part A*, 84, 57-65. <https://doi.org/10.1016/j.shpsa.2020.08.002>
- APA DSM Team. (2021). *Guide to Submitting Proposals for Changes to DSM-5*. American Psychiatric Association. Consultado el 13 de enero de 2025 de <https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/DSM5-Proposal-Submissions-General-Guidance.pdf>
- Berridge, K. C. (2009). Wanting and Liking: Observations from the Neuroscience and Psychology Laboratory. *Inquiry*, 52(4), 378-398. <https://doi.org/10.1080/00201740903087359>
- Boorse, C. (2014). A Second Rebuttal on Health. *Journal of Medicine and Philosophy*, 39(6), 683-724. <https://doi.org/10.1093/jmp/jhu035>
- Burdman, F. (2024). Recalcitrant Desires in Addiction. En S. Amaya, D. Shoemaker y M. Vargas (eds.), *Oxford Studies in Agency and Responsibility: Volume 8. Non-Ideal Agency and Responsibility* (pp. 58-80). Oxford University Press. <https://doi.org/https://doi.org/10.1093/oso/9780198910114.003.0004>
- Burdman, F. (2025). On the Grounds for Calling Addiction a Disease. *Análisis Filosófico*, 45(1), 203-231. <https://doi.org/10.36446/af.e1048>

¹⁶ El trabajo de FB fue apoyado por el proyecto Fondecyt Iniciación N° 11250503 (ANID, Chile). El trabajo de JRL fue apoyado por el proyecto Fondecyt Iniciación N° 11250401 (ANID, Chile). El trabajo de FP y de JRL fue apoyado por el proyecto Fondecyt Regular N° 1250205 (ANID, Chile).

- Butlin, P. y Papineau, D. (2017). Normal and Addictive Desires. En N. Heather y G. Segal (eds.), *Addiction and Choice: Rethinking the Relationship* (pp. 99-115). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198727224.003.0006>
- Carrier, M. (2011). Underdetermination as an Epistemological Test Tube: Expounding Hidden Values of the Scientific Community. *Synthese*, 180(2), 189-204. <https://doi.org/10.1007/s11229-009-9597-6>
- Colombo, M. (2014). Deep and Beautiful: The Reward Prediction Error Hypothesis of Dopamine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 45, 57-67. <https://doi.org/10.1016/j.shpsc.2013.10.006>
- Cooper, R. (2004). What is Wrong with the DSM? *History of Psychiatry*, 15(1), 5-25. <https://doi.org/10.1177/0957154X04039343>
- Cosmides, L. y Tooby, J. (1999). Toward an Evolutionary Taxonomy of Treatable Conditions. *Journal of Abnormal Psychology*, 108(3), 453-464. <https://doi.org/10.1037/0021-843X.108.3.453>
- Cummins, R. (1975). Functional Analysis. *The Journal of Philosophy*, 72(20), 741-765. <https://doi.org/10.2307/2024640>
- De Haro, S. (2023). The Empirical Under-Determination Argument Against Scientific Realism for Dual Theories. *Erkenntnis*, 88, 117-145. <https://doi.org/10.1007/s10670-020-00342-0>
- Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. Bradford Books.
- Dennett, D. C. (1987). *The Intentional Stance*. The MIT Press.
- Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy*, 88(1), 27-51. <https://doi.org/10.2307/2027085>
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster.
- Dennett, D. C. (2017). Darwin and the Overdue Demise of Essentialism. En D. L. Smith (ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism* (pp. 9-22). Cambridge University Press. <https://doi.org/10.1017/9781107295490.002>
- Dill, B. y Holton, R. (2014). The Addict in Us all. *Frontiers in Psychiatry*, 5. <https://doi.org/10.3389/fpsy.2014.00139>
- Durrant, R., Adamson, S., Todd, F. y Sellman, D. (2009). Drug Use and Addiction: An Evolutionary Perspective. *Australian and New Zealand Journal of Psychiatry*, 43(11), 1049-1056. <https://doi.org/10.3109/00048670903270449>

- Flanagan, O. (2025). *What Is It Like to Be an Addict?* Oxford University Press. <https://doi.org/10.1093/oso/9780199388929.001.0001>
- Fodor, J. A. (2000). *The Mind Doesn't Work that Way: The Scope and Limits of Computational Psychology*. The MIT press. <https://doi.org/10.7551/mitpress/4627.001.0001>
- Garson, J. (2019). *What Biological Functions Are and Why They Matter*. Cambridge University Press. <https://doi.org/10.1017/9781108560764>
- Ginnobili, S. y Roffé, A. (2017). Dos usos de los modelos de optimalidad en las explicaciones por selección natural. *Metatheoria*, 8(1), 43-55. <https://doi.org/10.48160/18532330me8.161>
- Gorelick, D. A., Zangen, A. y George, M. S. (2014). Transcranial Magnetic Stimulation in the Treatment of Substance Addiction. *Annals of the New York Academy of Sciences*, 1327(1), 79-93. <https://doi.org/10.1111/nyas.12479>
- Grace, A. A., Floresco, S. B., Goto, Y. y Lodge, D. J. (2007). Regulation of Firing of Dopaminergic Neurons and Control of Goal-Directed Behaviors. *Trends in Neurosciences*, 30(5), 220-227. <https://doi.org/10.1016/j.tins.2007.03.003>
- Harding, S. (1986). *The Science Question in Feminism*. Cornell University Press.
- Heilig, M., MacKillop, J., Martinez, D., Rehm, J., Leggio, L. y Vanderschuren, L. J. M. J. (2021). Addiction as a Brain Disease Revised: Why It Still Matters, and the Need for Consilience. *Neuropsychopharmacology*, 46(10), 1715-1723. <https://doi.org/10.1038/s41386-020-00950-y>
- Holton, R. y Berridge, K. (2013). Addiction between Compulsion and Choice. En N. Levy (ed.), *Addiction and Self-Control: Perspectives from Philosophy, Psychology, and Neuroscience* (pp.239-268). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199862580.003.0012>
- Kostko, A. (2019). Inductive Risks and Psychiatric Classification. En R. Bluhm y Ş. Tekin (eds.), *The Bloomsbury Companion to Philosophy of Psychiatry* (pp. 197-215). Bloomsbury Academic. <https://doi.org/10.5040/9781350024090.ch-010>
- Loaiza, J. R. (2024). Functionalism and the Emotions. *The British Journal for the Philosophy of Science*, 75(1), 233-251. <https://doi.org/10.1086/715207>
- Longino, H. E. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>
- Millikan, R. (1984). *Language, Thought and Other Biological Categories*. The MIT Press. <https://doi.org/10.7551/mitpress/4124.001.0001>

- Montague, P. R., Dayan, P. y Sejnowski, T. J. (1996). A Framework for Mesencephalic Dopamine Systems Based on Predictive Hebbian Learning. *Journal of Neuroscience*, 16(5), 1936-1947. <https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996>
- Murphy, D. (2009). Psychiatry and the Concept of Disease as Pathology. En M. Broome y L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical perspectives* (pp. 103-118). Oxford University Press. <https://doi.org/10.1093/med/9780199238033.003.0006>
- Murphy, D. y Stich, S. (2000). Darwin in the Madhouse: Evolutionary Psychology and the Classification of Mental Disorders. En P. Carruthers y A. Chamberlain (eds.), *Evolution and the Human Mind: Modularity, Language and Meta-Cognition* (pp. 62-92). Cambridge University Press. <https://doi.org/10.1017/CBO9780511611926.005>
- Nesse, R. M. y Berridge, K. C. (1997). Psychoactive Drug Use in Evolutionary Perspective. *Science*, 278(5335), 63-66. <https://doi.org/10.1126/science.278.5335.63>
- Nestler, E. J. (2005). Is There a Common Molecular Pathway for Addiction? *Nature Neuroscience*, 8(11). <https://doi.org/10.1038/nn1578>
- Papineau, D. (2017). Teleosemantics. En D. L. Smith (ed.), *How Biology Shapes Philosophy: New Foundations for Naturalism* (pp. 95-120). Cambridge University Press. <https://doi.org/10.1017/9781107295490.006>
- Quine, W. V. O. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60(1), 20-43. <https://doi.org/10.2307/2181906>
- Quine, W. V. O. (1975). On Empirically Equivalent Systems of the World. *Erkenntnis*, 9(3), 313-328. <https://doi.org/10.1007/BF00178004>
- Richters, J. E. y Hinshaw, S. P. (1999). The Abduction of Disorder in Psychiatry. *Journal of Abnormal Psychology*, 108(3), 438-445. <https://doi.org/10.1037/0021-843X.108.3.438>
- Robinson, T. y Berridge, K. (2008). The Incentive Sensitization Theory of Addiction: Some Current Issues. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1507), 3137-3146. <https://doi.org/10.1098/rstb.2008.0093>
- Ross, D. (2020). Addiction Is Socially Engineered Exploitation of Natural Biological Vulnerability. *Behavioural Brain Research*, 386, 112598. <https://doi.org/10.1016/j.bbr.2020.112598>
- Schroeder, T. (2004). *Three Faces of Desire*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195172379.001.0001>

- Schultz, W. (2016). Dopamine Reward Prediction Error Coding. *Dialogues in Clinical Neuroscience*, 18(1), 23-32. <https://doi.org/10.31887/DCNS.2016.18.1/wschultz>
- Tooby, J. y Cosmides, L. (1990). The Past Explains the Present. *Ethology and Sociobiology*, 11(4-5), 375-424. [https://doi.org/10.1016/0162-3095\(90\)90017-Z](https://doi.org/10.1016/0162-3095(90)90017-Z)
- Verhulst, B., Neale, M. C. y Kendler, K. S. (2015). The Heritability of Alcohol Use Disorders: A Meta-Analysis of Twin and Adoption Studies. *Psychological Medicine*, 45(5), 1061-1072. <https://doi.org/10.1017/S0033291714002165>
- Volkow, N.D., Koob, G.F. y McLellan, A.T. (2016). Neurobiologic Advances from the Brain Disease Model of Addiction. *New England Journal of Medicine*, 374(4), 363-371. <https://doi.org/10.1056/nejmra1511480>
- Wakefield, J. C. (1992). The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values. *American Psychologist*, 47(3), 373-388. <https://doi.org/10.1037/0003-066X.47.3.373>
- Wakefield, J. C. (2017a). Addiction and the Concept of Disorder, Part 1: Why Addiction is a Medical Disorder. *Neuroethics*, 10(1), 39-53. <https://doi.org/10.1007/s12152-016-9300-9>
- Wakefield, J. C. (2017b). Addiction and the Concept of Disorder, Part 2: Is Every Mental Disorder a Brain Disorder? *Neuroethics*, 10(1), 55-67. <https://doi.org/10.1007/s12152-016-9301-8>
- Wakefield, J. C. (2020). Addiction from the Harmful Dysfunction Perspective: How There Can Be a Mental Disorder in a Normal Brain. *Behavioural Brain Research*, 389, 112665. <https://doi.org/10.1016/j.bbr.2020.112665>
- Zawidzki, T. (2007). *Dennett*. Oneworld Publications.

